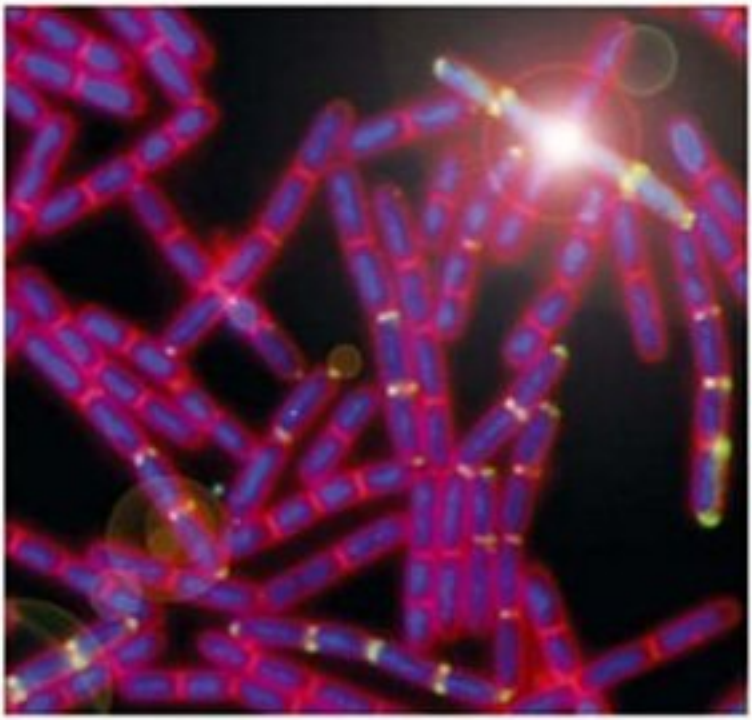
Wolfgang Schumann

WILEY-VCH

Dynamics of the Bacterial Chromosome

Structure and Function



VISIT…

Wolfgang Schumann

Dynamics of the Bacterial Chromosome

Dynamics of the Bacterial Chromosome. Structure and Function. W. Schumann
Copyright © 2006 Wiley-VCH Verlag GmbH & Co. KGaA
ISBN 3-527-30496-7

Related Titles

Sensen, C. W. (ed.)

Handbook of Genome Research

Genomics, Proteomics, Metabolomics, Bioinformatics, Ethical and Legal Issues

2005

ISBN 3-527-31348-6

Dale, J. W., Park, S.

Molecular Genetics of Bacteria

2004

ISBN 0-470-85084-1

Reece, R. J.

Analysis of Genes and Genomes

2004

ISBN 0-470-84379-9

Gellissen, G. (ed.)

Hansenula polymorpha

Biology and Applications

2002

ISBN 3-527-30341-3

Wolfgang Schumann

Dynamics of the Bacterial Chromosome

Structure and Function

WILEY-VCH Verlag GmbH & Co. KGaA

The Author

Dr. Wolfgang Schumann
Institute of Genetics
University of Bayreuth
95440 Bayreuth
Germany

Cover illustration:
The picture shows DAPI (blue) and FM4-64-stained (red) pre-divisional sporangia of *Bacillus subtilis* of a RacA-GFP-producing strain at h 1.5 of sporulation. RacA-GFP (green) is seen as foci at the poles and as a haze over the nucleoids. From the work of S. Ben-Yehuda, D. Rudner and R. Losick **2003**, Science 299, 532.

All books published by Wiley-VCH are carefully produced. Nevertheless, authors, editors, and publisher do not warrant the information contained in these books, including this book, to be free of errors. Readers are advised to keep in mind that statements, data, illustrations, procedural details or other items may inadvertently be inaccurate.

Library of Congress Card No.:
applied for

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library.

Bibliographic information published by Die Deutsche Bibliothek
Die Deutsche Bibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data is available in the Internet at <http://dnb.ddb.de>.

© 2006 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

All rights reserved (including those of translation into other languages). No part of this book may be reproduced in any form – by photoprinting, microfilm, or any other means – nor transmitted or translated into a machine language without written permission from the publishers. Registered names, trademarks, etc. used in this book, even when not specifically marked as such, are not to be considered unprotected by law.

Typesetting Fotosatz Detzner, Speyer
Printing Strauss GmbH, Mörlenbach
Bookbinding J. Schäffer GmbH, Grünstadt

Printed in the Federal Republic of Germany
Printed on acid-free paper

ISBN-13: 978-3-527-30496-7
ISBN-10: 3-527-30496-7

Contents

Foreword *XI*

Bacterial Species and their Abbreviation *XIII*

Color Plates *XV*

1 Structure of the Bacterial Cell *1*
1.1 The Cytoplasm Compartment *1*
1.2 The Cytoplasmic Membrane Compartment *3*
1.3 The Cell Wall Compartment *8*
1.4 The Outer Membrane Compartment *9*
1.5 The Periplasmic Compartment *15*
1.6 Extracellular Matrices *19*
1.7 Appendages *21*

2 Organization of the Bacterial Chromosome *29*
2.1 Structure of the Chromosomes *29*
2.2 Principles to Compact the Bacterial Chromosome *31*
2.2.1 Superhelicity *31*
2.2.2 Histone-like Proteins in *E. coli* *32*
2.3 Organization of the Bacterial Chromosome into Genes and Repetitive Sequences *36*
2.3.1 Genes *36*
2.3.2 Repetitive Sequences *37*
2.4 Large Rearrangements Within the Chromosome *44*
2.4.1 Duplications *44*
2.4.2 Deletions *45*
2.4.3 Inversions *49*

3 The Bacterial Cell Cycle: Replication of the Chromosome, Partitioning and Cell Division *51*
3.1 Replication *51*
3.1.1 Replication of Circular Chromosomes *51*

Dynamics of the Bacterial Chromosome. Structure and Function. W. Schumann
Copyright © 2006 Wiley-VCH Verlag GmbH & Co. KGaA
ISBN 3-527-30496-7

3.1.2 Replication of Linear Chromosomes and Plasmids *67*
3.2 Partitioning (Segregation) of the Daughter Chromosomes *68*
3.3 Cell Division *73*
3.4 Plasmid- and Chromosome-encoded Toxin–Antitoxin Modules *79*
3.4.1 Postsegregational Killing of Plasmid-free Cells *80*
3.4.2 Chromosomal Toxin–Antitoxins *81*

4 Recombination *83*
4.1 Homologous Recombination *83*
4.1.1 The RecBCD Pathway of Homologous Recombination *85*
4.1.2 The RecF Pathway of Homologous Recombination *90*
4.1.3 Additional Homologous Recombination Functions *91*
4.1.4 Recovery of Replication at a Blocking DNA Lesion *92*
4.2 Sequence-specific Recombination *97*
4.2.1 Integration/Excision Systems *99*
4.2.2 Resolvase Systems *100*
4.2.3 Inversion Systems *103*
4.2.4 Shufflons *107*
4.2.5 Integrons *107*
4.2.6 Homing Endonucleases *110*
4.3 Illegitimate Recombination *112*
4.3.1 Insertion Sequence Elements *113*
4.3.2 Transposons *114*
4.3.3 Transposing Bacteriophages *115*
4.3.4 Mechanisms of Transposition *116*
4.3.5 Conjugative and Integrative Elements *118*
4.3.6 Nonhomologous Endjoining (NHEJ) *119*

5 Origin of Mutations and Repair of DNA Lesions *121*
5.1 Classes of Mutations *121*
5.2 Origin of Mutations *124*
5.2.1 Spontaneous Mutations *124*
5.2.2 Induced Mutations *126*
5.2.3 Detection Systems for Mutations *128*
5.2.4 Adaptive Mutations *129*
5.3 Repair of DNA Lesions *131*
5.3.1 The Methyl-mediated Mismatch Repair System *132*
5.3.2 Very Short Patch Repair *133*
5.3.3 The Nucleotide Excision Repair Systems *134*
5.3.4 Pathways of Base Excision Repair *136*
5.3.5 Repair of Oxidized Nucleotides *138*
5.3.6 Photoreactivation of Cyclobutane Dimers *140*
5.3.7 Transcription-coupled Nucleotide Excision Repair *141*
5.3.8 The Adaptive Response *141*
5.3.9 The SOS Response *143*
5.3.10 Replication and Repair *148*

5.3.11 Repair of Interstrand Crosslinks in DNA *150*
5.3.12 *Deinococcus radiodurans* *151*

6 Principles of Gene Regulation *153*
6.1 Regulation at the Level of DNA *153*
6.1.1 Alterations of the Structure of the DNA *154*
6.1.2 GATC Methylation and Gene Expression *157*
6.1.3 Programmed DNA Rearrangements Within the Chromosome *159*
6.1.4 Transcriptional Silencing and Cryptic Genes *159*
6.2 Regulation at the Level of Transcription *162*
6.2.1 The DNA-dependent RNA Polymerase *162*
6.2.2 Regions of the Promoter *168*
6.2.3 Initiation of Transcription *170*
6.2.4 Elongation of Transcription *173*
6.2.5 Termination of Transcription *175*
6.2.6 Anti-sigma Factors *177*
6.2.7 Anti-termination/Attenuation of Transcription *179*
6.2.8 Regulators of the DNA-dependent RNA Polymerase *188*
6.3 Post-transcriptional Regulation *220*
6.3.1 Protein-mediated Regulation *220*
6.3.2 Riboregulation *222*
6.3.3 Riboswitches *227*
6.3.4 mRNA Decay *230*
6.3.5 Introns *239*
6.3.6 RNA-binding Proteins *242*
6.4 Regulation at the Level of Translation *244*
6.4.1 Initiation of Translation *247*
6.4.2 Elongation of Translation *248*
6.4.3 Termination of Translation *252*
6.4.4 Ribosome Recycling *253*
6.4.5 Translational Control *255*
6.4.6 The 21. and 22. Amino Acid Selenocysteine and Pyrrolysine *258*
6.4.7 Mechanisms to Create Multiple Gene Products From a Single Cistron *259*
6.5 Post-translational Regulation *266*
6.5.1 Proteolysis *266*
6.5.2 Protein Splicing *266*
6.5.3 Protein Glycosylation *266*
6.5.4 Circular Polypeptides *268*

7 Protein Quality Control Through Molecular Chaperones, Protein Folding Catalysts and ATP-dependent Proteases *271*
7.1 Introduction *271*
7.2 Molecular Chaperones *271*
7.2.1 Folder Chaperones *272*

7.2.2 Holder Chaperones *279*
7.2.3 Disaggregating Chaperones *287*
7.2.4 Intramolecular Chaperones *289*
7.2.5 Folding of Nascent Polypeptide Chains *290*
7.3 Protein-folding Catalysts *292*
7.3.1 Peptidyl-prolyl Isomerases *292*
7.4 Disulfide Bond Formation *293*
7.5 Proteases *297*
7.5.1 ATP-dependent Proteases *298*
7.5.2 ATP-independent Proteases *305*

8 Secretion of Proteins *309*
8.1 Introduction *309*
8.2 Protein Sorting Signals for Subcellular Localization *310*
8.3 Cytosolic Targeting Factors *315*
8.4 Translocases Move Newly Made Proteins Into or Across Membranes *316*
8.4.1 The Sec Pathway of Protein Translocation Through the Inner Membrane *317*
8.4.2 The Tat Pathway of Secretion Proteins Through the Inner Membrane *321*
8.4.3 Insertion Proteins Into the Inner Membrane *323*
8.5 Anchoring of Lipoproteins Into the Outer Membrane *324*
8.6 Protein Translocation Pathways Through Both Membranes of Gram-negative Bacteria *326*
8.6.1 The Type I Secretion Pathway *326*
8.6.2 The Type II Secretion Pathway *327*
8.6.3 The Type III Secretion Pathway *330*
8.6.4 The Type IV Secretion Pathway *333*
8.6.5 Type V Secretion Pathway (Autotransporter) *335*
8.6.6 Vesicle-mediated Export *336*
8.7 Cell Wall-anchored Surface Proteins *337*
8.8 S-Layers *339*

9 Stress Genes and Their Regulation *341*
9.1 The Stress Response Pathway *341*
9.2 The Heat Shock Response *342*
9.3 The Cold Shock Response *349*
9.4 The Osmotic Shock Response *353*
9.4.1 The Hyperosmotic Shock Response *354*
9.4.2 The Hypoosmotic Shock Response *359*
9.5 The Oxidative Shock Response *361*
9.5.1 The OxyR Regulon *361*
9.5.2 The SoxRS Regulon *362*
9.6 pH-regulated Genes *364*

9.6.1 Acid Stress *367*
9.6.2 Alkaline Stress *369*
9.7 Metabolic Stress: the Stringent Response *371*
9.8 Nutrient Limitations (Starvation Stress) *373*
9.9 Envelope Stress Response *376*

10 Exchange of Genes: Transformation, Conjugation and Transduction *381*
10.1 Transformation *381*
10.1.1 Transformation of Gram-positive Bacteria *384*
10.1.2 Transformation of Gram-negative Bacteria *387*
10.2 Conjugation *388*
10.2.1 Gram-negative Bacteria *389*
10.2.2 Gram-positive Bacteria *396*
10.2.3 Conjugative Transposons *398*
10.3 Transduction *398*
10.3.1 Generalized Transduction *400*
10.3.2 Specialized Transduction *401*
10.4 Horizontal Gene Transfer *402*

Glossary *405*

Subject Index *413*

Foreword

Over the past 50 years, molecular microbiology has provided fundamental contributions not only to the understanding of microbiology as a unique discipline but also to the understanding of what makes a living cell. In 1945 Schrödinger in his famous lecture ''What is life?'' addressed the challenging task of defining the properties of a necessary (e.g. minimal/required) information-containing system. More than half a century later, we can state that many aspects of this fundamental question ''What is life?'' have been answered. This series of great discoveries started with the elucidation of the DNA structure by F. Crick and J. Watson more than 50 years ago and was subsequently expanded by the exploration of the basic mechanisms of DNA replication and gene expression. These scientific achievements were complemented and fostered by the development of crucial techniques, such as gene cloning, DNA sequencing and PCR. The discoveries culminated in an event that will certainly have a fixed position in future text books of science history: the first complete elucidation of a genome sequence of a living organism, *Haemophilus influenzae*, in 1995. At the beginning of 2006, the complete genomes of nearly 300 bacterial species are available. Knowledge of the complete genome sequence is an essential prerequisite in order to gain a comprehensive understanding of the molecular mechanisms of life. The genome sequence, however, only provides the ''blueprint'' of life, not life itself. Now, functional genomics such as transcriptomics, proteomics, metabolomics and bioinformatics are required to bring the blueprint of life to the real life of living organisms. The combination of the expertise and accumulated knowledge in the traditional disciplines such as microbiology, biochemistry, molecular genetics and molecular biology with this panoramic view of the recent ''omics technologies'' on the cell as an entity will enable/facilitate a new quality in the understanding of what makes a cell viable. Because of their low complexity, single-cell bacterial systems such as *Escherichia coli*, *Bacillus subtilis* and others constitute perfect model systems for tackling such an ambitious goal as understanding life as an entity. The unexpected finding that, even in the best analyzed model organisms such as *Escherichia coli*, one-third of all genes codes for proteins with still unknown functions emphasizes the challenge and indicates that many pages of the ''bible of life'' are still empty.

There is a comprehensive literature on the entire field of molecular microbiology, from the structure, flexibility and stability of bacterial genomes to the various facets of the regulation of gene expression, a field that even in bacteria includes mechan-

Dynamics of the Bacterial Chromosome. Structure and Function. W. Schumann
Copyright © 2006 Wiley-VCH Verlag GmbH & Co. KGaA
ISBN 3-527-30496-7

isms of not only transcriptional initiation but also transcriptional elongation and termination and regulation at the posttranscriptional, translational and posttranslational level. Even if the first description of the regulation of a bacterial operon, the *lac*-operon in *E. coli* by Jacob and Monod, was a milestone in the history of microbiology, honored by the Nobel Price in 1965, we had however to realize that this regulation mechanism represents only one of hundreds of mechanisms evolved during the three billion years that bacteria have populated our planet. For those involved in teaching students, it is difficult to keep up with the pace of the development of all these different fields of molecular microbiology. A textbook covering all these aspects of molecular microbiology is equally useful for teachers and students in microbiology at universities, because this very genuine, innovative and essential new field of microbiology needs more space in the teaching programs for life science students.

Wolfgang Schumann, professor for genetics at the University Bayreuth, accepted the challenge of writing a textbook on molecular bacteriology. For many years, he has lectured students of biology and biochemistry on the genetics and molecular biology of bacteria. He is appreciated as an expert in molecular genetics and bacterial gene regulation and his contributions to many fields of molecular genetics, e.g. the mechanisms of heat induction in Gram-positive bacteria and the discovery of the CIRCE element (to mention only some of them), are accepted worldwide.

I have always been impressed by his detailed knowledge of the scientific literature on quite different fields of molecular microbiology. This is one reason which convinced me that Wolfgang Schumann is the right person to take over this ambitious task. I have already used the chance to read some of the chapters of this book in preparing my lectures at the University of Greifswald this year. I enjoyed this reading, very conveniently providing me with an excellent survey. Assessing the recent literature critically and in detail an essential part of the training of young students has been done in an excellent way. All essential aspects of molecular bacteriology (which might be extended by a similar textbook by other experts on the molecular microbiology of Archaea or eukaryotic microorganisms) have been addressed in a very authentic, comprehensive manner and have been illustrated by a lot of impressive and useful figures.

This book covers the most essential chapters in molecular microbiology, starting with the structure of the bacterial cell via the organization of bacterial chromosomes to the bacterial cell cycle, followed by chapters on molecular genetics, recombination and mutations and their repair. A comprehensive chapter deals with the various mechanisms of gene regulation, followed by a description of the role of chaperones in protein quality control and protein secretion mechanisms. An extra chapter on stress genes and their regulation is a reflection of the specific research interests of the author. A final chapter on gene transfer makes the book complete. I just missed a chapter on the prokaryotic development of bacteria, including endospore-forming bacteria such as *B. subtilis*, heterocyst formation in *Anabaena* and the dimorphic life cycle of *Caulobacter crescentus*, to mention some.

All in all, this new book offers an actual, comprehensive view of the present state of the art in the field of molecular bacteriology. It is a very valuable source of information not only for students, but also for scientists who wish to become acquainted with one of the most exciting fields of current microbiology.

Greifswald, 04.04.2006 Michael Hecker

Bacterial Species and their Abbreviations

A. actinomycetemcomitans	*Actinobacillus actinomycetemcomitans*
A. salmonicida	*Aeromonas salmonicida*
A. tumefaciens	*Agrobacterium tumefaciens*
A. vinelandii	*Azotobacter vinelandii*
B. bronchiseptica	*Bordetella bronchiseptica*
B. burgdorferi	*Borrelia burgdorferi*
B. japonicum	*Bradyrhizobium japonicum*
B. megaterium	*Bacillus megaterium*
B. pertussis	*Bordetella pertussis*
B. polymyxa	*Bacillus polymyxa*
B. pseudomallei	*Burkholderia pseudomallei*
B. hyodysenteriae	*Brachyspira hyodysenteriae*
B. subtilis	*Bacillus subtilis*
C. acetobutylicum	*Clostridium acetobutylicum*
C. crescentus	*Caulobacter crescentus*
C. difficile	*Clostridium difficile*
C. diphtheriae	*Corynebacterium diphtheria*
C. freundii	*Citrobacter freundii*
C. jejuni	*Campylobacter jejuni*
C. perfingens	*Clostridium perfringens*
C. psittacii	*Chlamydia psittacii*
C. trachomatis	*Chlamydia trachomatis*
C. violaceum	*Chromobacterium violaceum*
D. radiodurans	*Deinococcus radiodurans*
E. amylovora	*Erwinia amylovora*
E. carotovora	*Erwinia carotovora*
E. cloacae	*Enterobacter cloacae*
E. coli	*Escherichia coli*
E. crysanthemie	*Erwinia crysanthemie*
E. faecalis	*Enterococcus faecalis*
H. ducreyi	*Haemophilus ducreyi*
H. halobium	*Halobacterium halobium*
H. influenzae	*Haemophilus influenzae*
H. pylori	*Helicobacter pylori*
H. salinarium	*Halobacterium salinarium*
K. oxytoca	*Klebsiella oxytoca*

Dynamics of the Bacterial Chromosome. Structure and Function. W. Schumann
Copyright © 2006 Wiley-VCH Verlag GmbH & Co. KGaA
ISBN 3-527-30496-7

K. pneumoniae	*Klebsiella pneumoniae*
L. lactis	*Lactococcus lactis*
L. monocytogenes	*Listeria monocytogenes*
M. genitalium	*Mycoplasma genitalium*
M. janashii	*Methanococcus janashii*
M. leprae	*Mycobacterium leprae*
M. loti	*Mesorhizobium loti*
M. pulmonis	*Mycoplasma pulmonis*
M. tuberculosis	*Mycobacterium tuberculosis*
M. xanthus	*Myxococcos xanthus*
N. gonorrhoeae	*Neisseria gonorrhoeae*
N. meningitis	*Neisseria meningitis*
P. aeruginosa	*Pseudomonas aeruginosa*
P. fluorescens	*Pseudomonas fluorescens*
P. gingivalis	*Porphyromonas gingivalis*
P. putida	*Pseudomonas putida*
P. syringae	*Pseudomonas syringae*
R. capsulatus	*Rhodobacter capsulatus*
R. etli	*Rhizobium etli*
R. solanacearum	*Ralstonia solanacearum*
S. agalactiae	*Streptococcus agalactiae*
S. albus	*Streptomyces albus*
S. ambofaciens	*Streptomyces ambofaciens*
S. aurantiaca	*Stigmatella aurantiaca*
S. aureus	*Staphylococcus aureus*
S. avermitilis	*Streptomyces avermitilis*
S. coelicor	*Streptomyces coelicor*
S. enterica	*Salmonella enterica*
S. flexneri	*Shigella flexneri*
S. hygroscopius	*Streptomyces hygroscopius*
S. meliloti	*Sinrhizobium meliloti*
S. pneumoniae	*Streptococcus pneumoniae*
S. pyogenes	*Streptococcus pyogenes*
S. sonnei	*Shigella sonnei*
S. tsukubaensis	*Streptomyces tsukubaensis*
S. typhi	*Salmonella typhi*
S. typhimurium	*Salmonella typhimurium*
T. aquaticus	*Thermus aquaticus*
T. maritima	*Thermotoga maritima*
T. thermophilus	*Thermus thermophilus*
V. alginolyticus	*Vibrio alginolyticus*
V. cholerae	*Vibrio cholerae*
V. fischeri	*Vibrio fischeri*
V. harveyi	*Vibrio harveyi*
X. campestris	*Xanthomonas campestris*
Y. enterocolitica	*Yersinia enterocolitica*
Y. pestis	*Yersinia pestis*

Color Plates

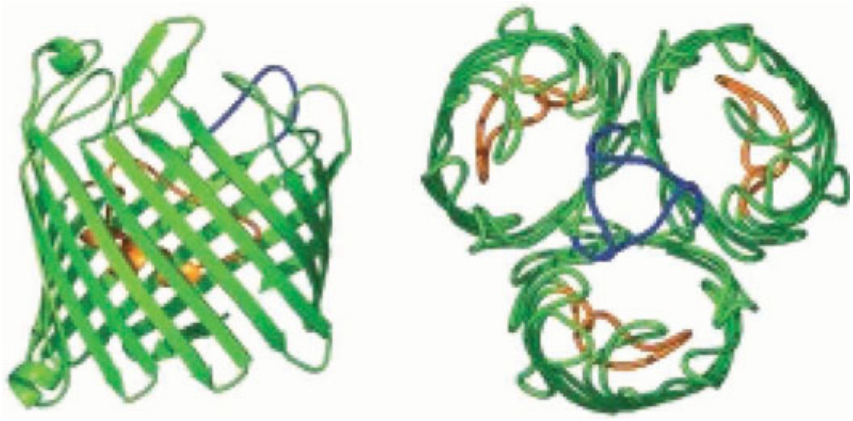


Fig. 1.2 Structure of the OmpF porin. View of the monomer from the side (left) and view of the trimer from the top. The loop inside the opening narrows the channel. H. Nikaido **2003**, *Microbiol. Mol. Biol. Rev.* 67, 593–556; Fig. 2.

Dynamics of the Bacterial Chromosome. Structure and Function. W. Schumann
Copyright © 2006 Wiley-VCH Verlag GmbH & Co. KGaA
ISBN 3-527-30496-7

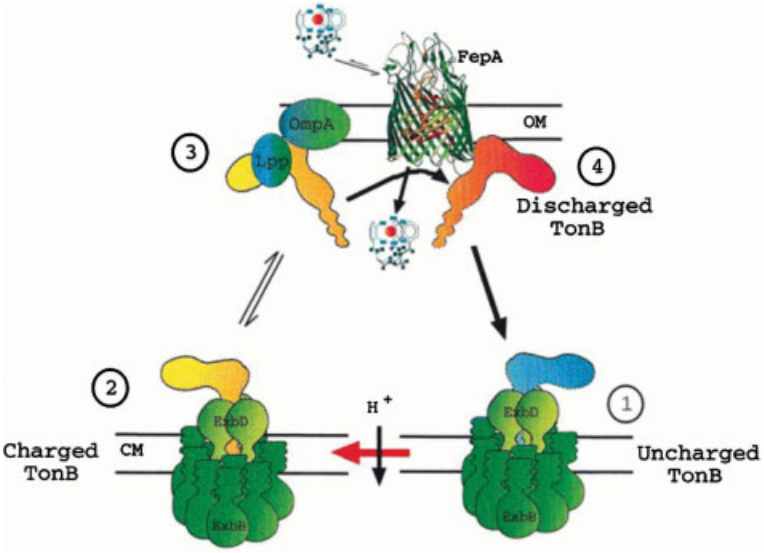


Fig. 1.3 Shuttle model for TonB-dependent energy transduction. Uncharged TonB (1) is energized by uptake of a proton (2) and shuttles to the outer membrane where is docks to OmpA/Lpp waiting for a transporter such as FepA with a bound ligand. Upon interaction with the Ton box of the the transporter the conformational energy is transduced to the transporter (3) triggering uptake of the ligand into the periplasm (4). The discharged TonB shuttles back to the ExbBD complex to become recharged. K. Postle, R.J. Kadner **2003**, *Mol. Microbiol.* 49, 869–882; Fig. 2.

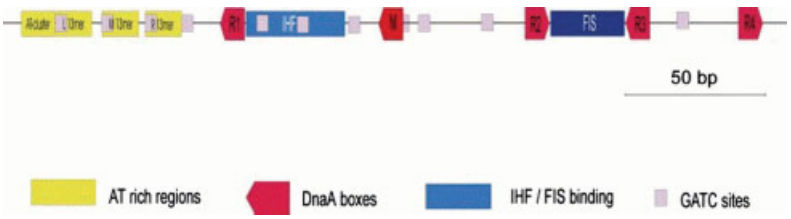


Fig. 3.1 Organization of the *oriC* region of *E. coli*. The *oriC* region of *E. coli* consists of five DnaA boxes called R1 through R4 and M, binding sites for the histone-like proteins FIS and IHF, three AT-rich regions and 11 GATC methylation sites. W. Messer **2002**, *FEMS Microbiol. Rev.* 26, 355; Fig. 1.

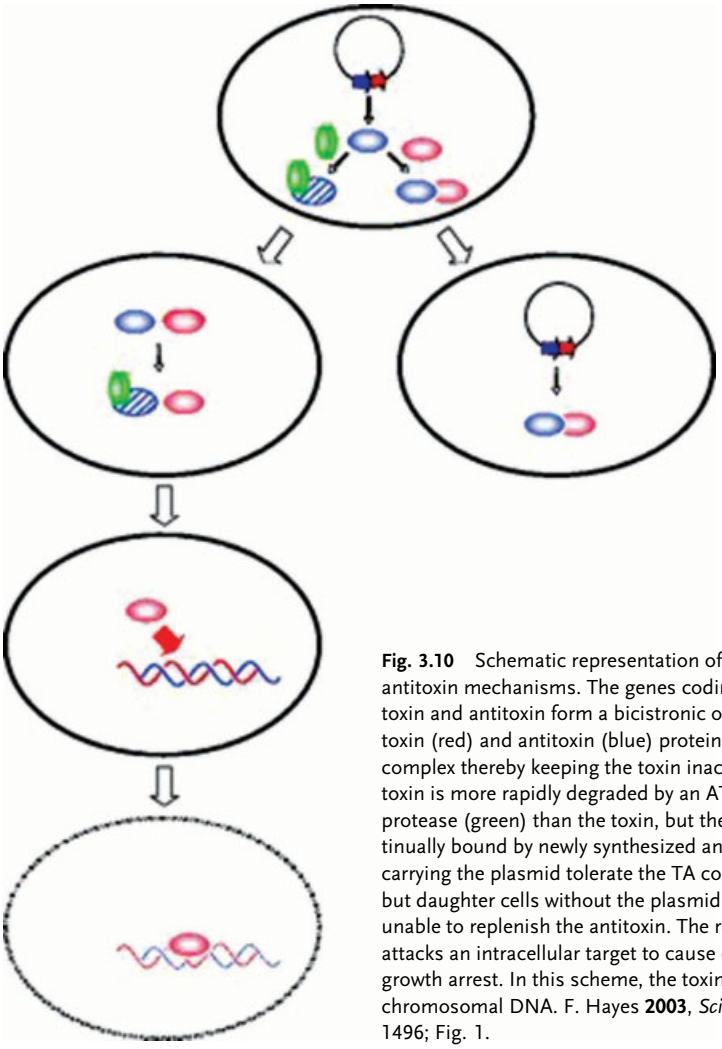


Fig. 3.10 Schematic representation of the toxin–antitoxin mechanisms. The genes coding for the toxin and antitoxin form a bicistronic operon. The toxin (red) and antitoxin (blue) proteins form a tight complex thereby keeping the toxin inactive. The antitoxin is more rapidly degraded by an ATP-dependent protease (green) than the toxin, but the latter is continually bound by newly synthesized antitoxin. Cells carrying the plasmid tolerate the TA complex (right), but daughter cells without the plasmid (left) are unable to replenish the antitoxin. The released toxin attacks an intracellular target to cause cell death or growth arrest. In this scheme, the toxin attacks the chromosomal DNA. F. Hayes **2003**, *Science* 301, 1496; Fig. 1.

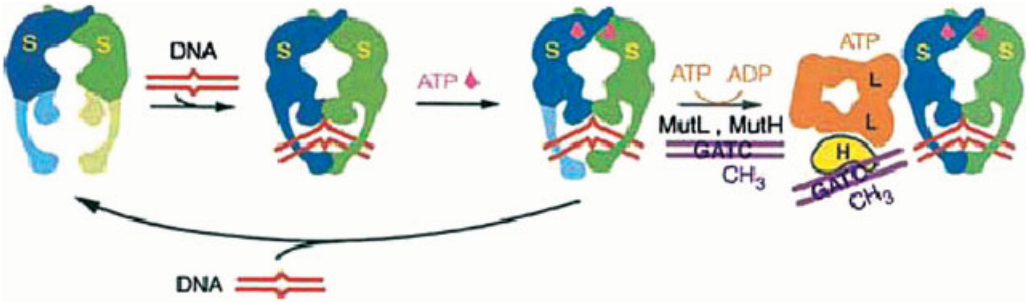


Fig. 5.2 Methyl-mediated mismatch repair. MutS recognizes and binds to the mismatch followed by binding of ATP. This complex recruits MutL which recognizes the nearest hemimethylated GATC and binds MutH. MutH cuts 5′ or 3′ to the unmodified GATC sequence, UvrD binds to the nick and unwinds the double-strand DNA followed by resynthesis and ligation. T.A. Kunkel, D.A. Erie **2005**, *Annu. Rev. Biochem.* 74, 681; Fig. 2C.

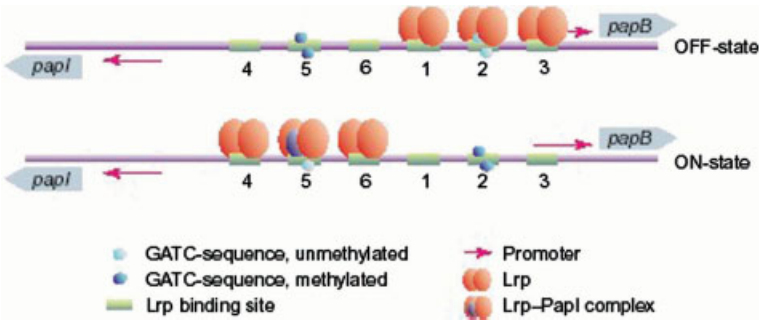


Fig. 6.3 Turning *pap* pilus synthesis on and off. See text for explanation who a cell can shift from the OFF- to the ON-state and vice versa. A. Løbner-Olesen, et al. **2005**, *Curr. Opin. Microbiol.* 8, 154–160.

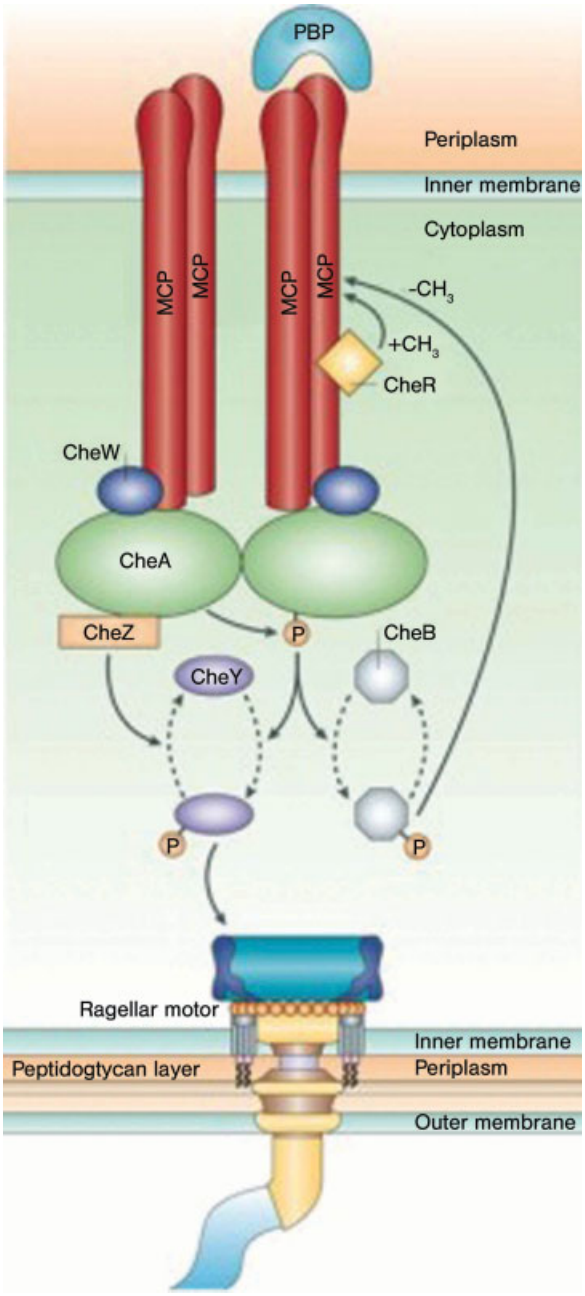


Fig. 6.27 Signalling pathways in bacterial chemotaxis. Dimeric MCPs form α-helical coiled-coil structures spanning the inner membrane. Chemical compounds (attractants and repellents) enter the periplasm through porins and interact either directly with the periplasmic domain of the appropriate chemoreceptor or indirectly through a *p*eriplasmatic *b*inding *p*rotein (PBP; MBP, RBP, GBP and DPP are *m*altose, *r*ibose, *g*alactose and *d*ipep*t*ide *b*inding *p*rotein, respectively). The CheW coupling molecule transduces the signals to the CheA sensor kinase which, after *trans*-phosphorylation, transfers the phosphate first to CheY and then to methylesterase CheB. CheY~P interacts with the flagellar motor to bring about a change in the direction. CheB~P competes with a constitutive methyltransferase, CheR, to control the degree of methylation of specific glutamates in the MCPs. This resets the signaling state of the chemoreceptors and allows them to adapt to the present concentration of attractant and to sense subsequent changes. Dephosphorylation of CheY~P is accelerated by the phosphatese CheZ. G.H. Wadhams **2004**, *Nat. Rev. Mol. Cell Biol.* 5, 1024–1037; Fig. 2.

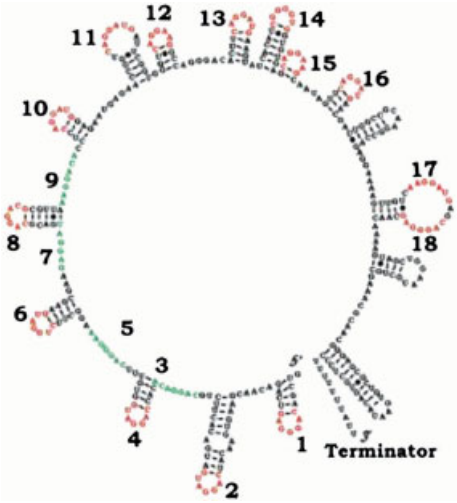


Fig. 6.37 Proposed structure of the CsrB RNA. This structure contains 18 repeated structures numbered 1 to 18 which may facilitate binding of CsrA. T. Romeo **1998**, *Mol. Microbiol.* 29, 1321–1330; Fig. 1.

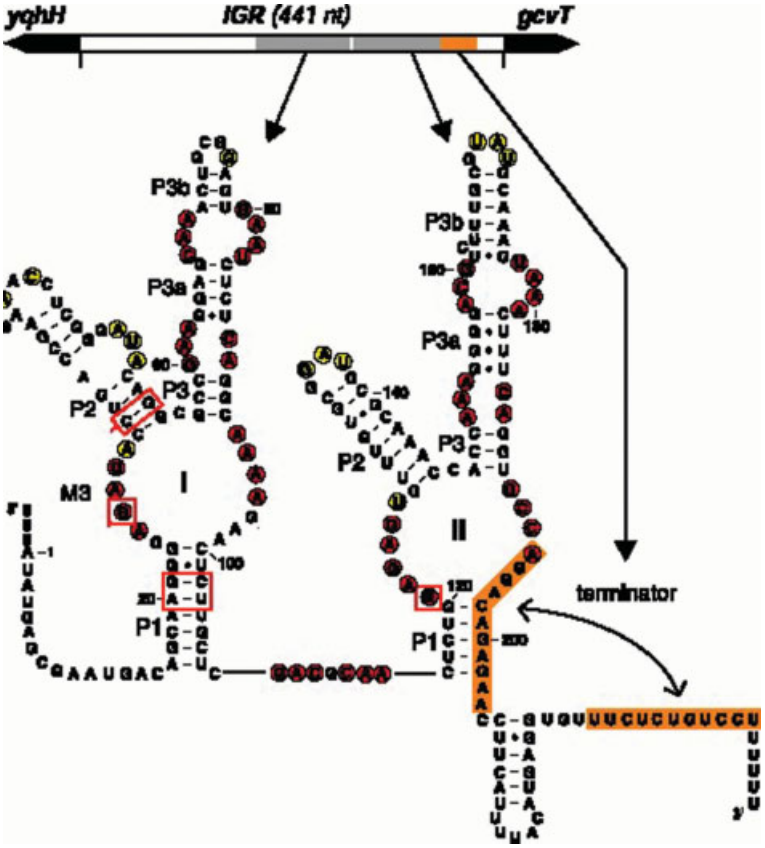


Fig. 6.38 The *B. subtilis gcvT* riboswitch. The intergenic region between *yqhH* and the *gcvPA-gcvPB* operon designated *gcvT* acts as a riboswitch and contains the two glycine sensing aptamers I and II. If the glycine concentration in the cell is low, the two aptamers remain unoccupied favoring formation of the terminator structure resulting in transcription attenuation. If the glycine concentration is high, it will bind to the two aptamers preventing formation of the transcription terminator, and the two structural genes will be expressed. M. Mandal, et al. **2004**, *Science* 306, 275; Fig. 4, modified.

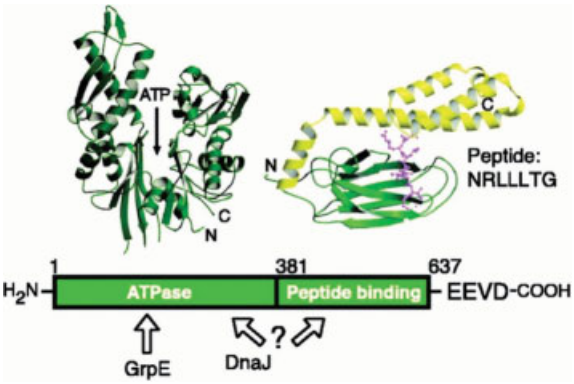


Fig. 7.1 Domain organization of the DnaK chaperone. DnaK consists of an N-terminal ATPase and a C-terminal peptide-binding domain. The ATPase domain consists of two lobes which form a deep cleft serving as a binding pocket for ATP. The peptide-binding domain (here shown with a heptapeptide substrate) forms a β sandwich composed of two sheets of four strands each, followed by α helices spanning back over the sandwich. Further indicated by open arrows are interaction sites for GrpE and DnaJ. F.U. Hartl, M. Hayer-Hartl **2002**, *Science* 295, 1852–1858; Fig. 3A.

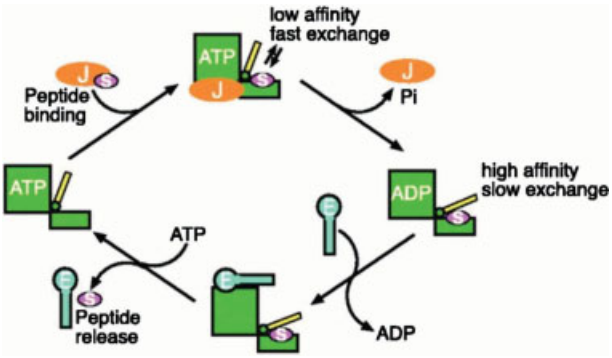


Fig. 7.2 Reaction cycle of the DnaK chaperone machine. DnaK-ATP with the open lid constitutes the active form able to accept non-native substrate polypeptide chains. These are either targeted to DnaK by DnaJ (as shown here) or bind directly to the peptide-binding domain of DnaK. The initial binding of the substrate protein is characterized by low affinity allowing fast dissociation. Binding of DnaJ to DnaK stimulates its ATPase activity resulting in the release of inorganic phosphate followed by closing of the lid, as shown in the cartoon. Now, the substrate protein is trapped and can start to refold. Later, GrpE binds to the complex, stimulates release of ADP followed by opening of the lid with concomitant dissociation of the substrate protein. After binding of ATP, a new reaction cycle can be started. F.U. Hartl, M. Hayer-Hartl **2002**, *Science* 295, 1852–1858; Fig. 3B.

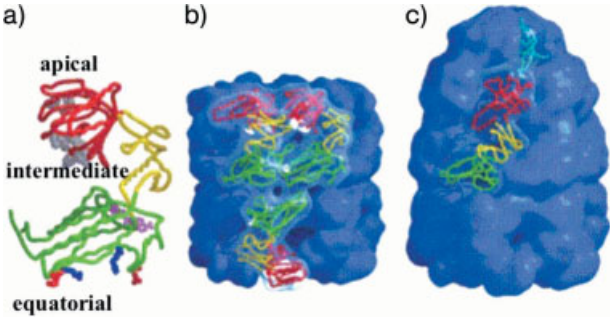


Fig. 7.3 The GroE chaperonin complex. (a) The three functional domains of the GroEL monomer. (b) Structure of two GroEL rings (14-mer) with three subunits shown at the front. (c) The GroEL-ADP-GroES complex with one GroEL and GroES subunit each. H.R. Saibil, N.A. Ranson **2002**, Trends Biochem. Sci. 27, 627–632; Fig. 2.

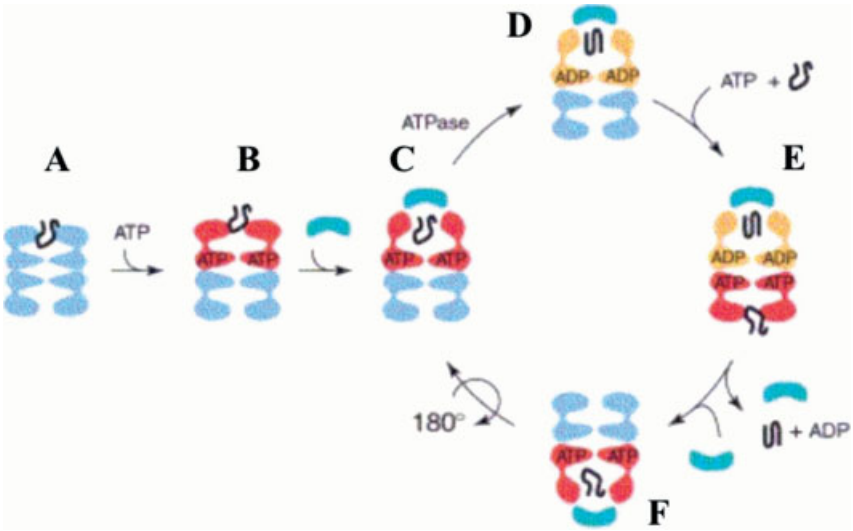


Fig. 7.4 The GroEL-GroES reaction cycle. (A) A substrate protein diffuses into the upper ring where it will bind to hydrophobic side chains exposed on the inner surface of the apical domain. (B) Next, each equatorial domain of the upper ring binds one molecule of ATP with positive cooperativity. (C) This in turn triggers binding of a GroES ring to the upper ring causing an upward movement of the apical domain, retraction of the hydrophobic side-chains and release of the substrate protein into the cavity of the upper GroEL ring where it starts to fold. (D) ATP is hydrolyzed and (E) another substrate molecule diffuses into the lower ring, followed by binding of ATP. (F) GroES diffuses away from the upper ring, followed by the substrate protein; and the substrate molecule bound to the lower ring gets encapsulated. H.R. Saibil, N.A. Ranson **2002**, Trends Biochem. Sci. 27, 627–632; Fig. 1.

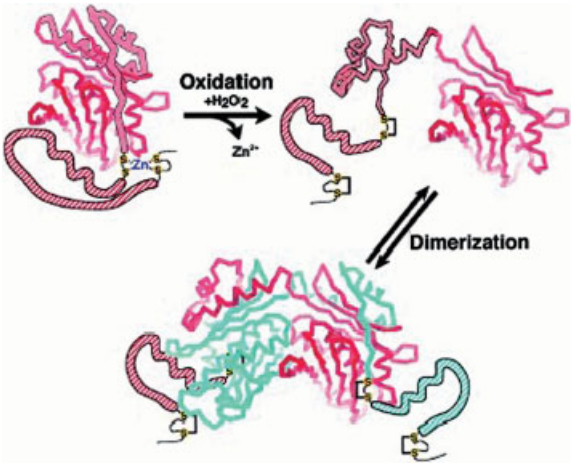


Fig. 7.8 Model of Hsp33 activation. Under reducing conditions, Hsp33 occurs in a monomeric form and four cysteine residues are involved in zinc coordination. After addition of H_2O_2, zinc is released and two intramolecular disulfide bonds are formed. Next, two oxidized Hsp33 monomers form a highly active dimer. P.C.F. Graf, U. Jacob **2002**, *Cell. Mol. Life Sci.* 59, 1624–1631; Fig. 1.

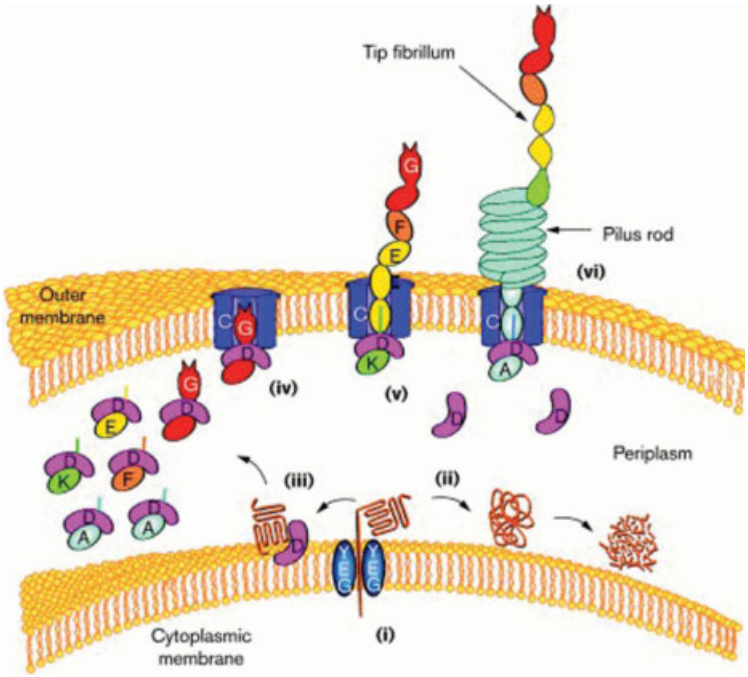


Fig. 7.9 Biogenesis of the P pilus. (i) The components of the P pilus and those involved in its biogenesis are synthesized in the cytoplasm and secreted into the periplasm using the Sec pathway. (ii) If components of the P pilus do not interact with the PapD chaperone, they fold inappropriately and are degraded by periplasmic proteases. (iii) Most P pilus components are bound by the PapD chaperone and guided to the PapC protein, an integral outer membrane protein acting as an usher (iv). (v) PapC monitors the correct assembly of the different components which form the active P pilus (vi). F.G Sauer, et al. **2000**, *Curr. Opin. Struct. Biol.* 10, 548–556; Fig. 1.

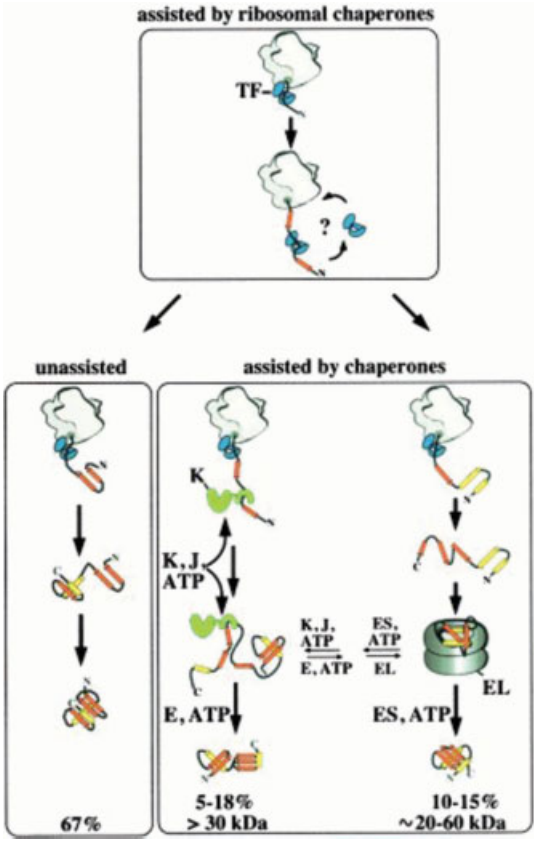


Fig. 7.13 Model for chaperone-assisted folding of nascent polypeptide chains in the cytoplasm. Nascent polypeptide chains first interact with the trigger factor (TF) loacted at the exit tunnel of the ribosome. While most proteins are able to fold in the complete absence of any folder chaperone, 5–18% are dependent on the DnaK and another 10–15% on the GroE team with overlapping specificity. B. Bukau, et al. **2000**, *Cell* 101, 119–122; Fig. 1.

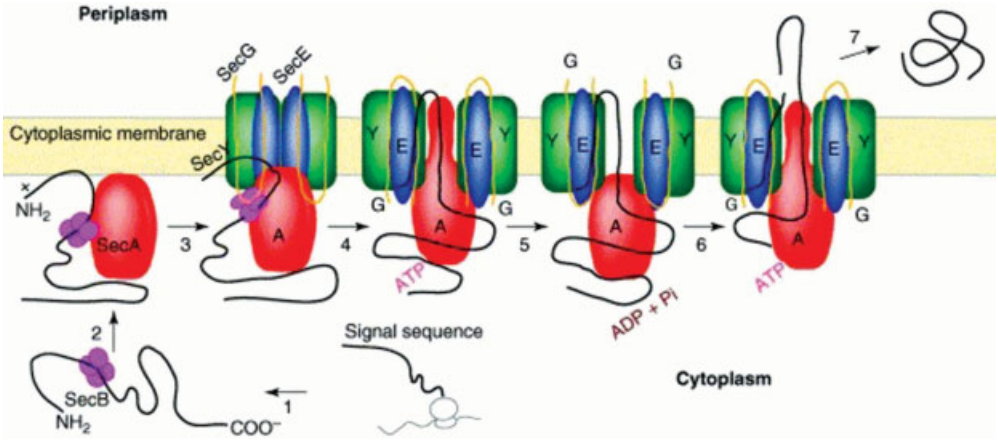


Fig. 8.5 The Sec or general translocation process. Most nascent polypeptide chains to be translocated are recognized and bound by the SecB chaperone (step 1) which target it to a SecA homodimer (step 2). Upon dissociation of SecB, SecA will bind to the inner membrane causing SecYEG to form the translocon (step 3). Next, SecA with the bound preprotein inserts into the channel (step 4) which requires bound ATP. Upon hydrolysis of ATP, SecA retracts (step 5), binds to another segment of the protein which is feeded together with SecA into the channel (step 6). During each translocation step, 20–30 amino acid residues are pushed through the translocon. H. Mori **2002**, *Trends Microbiol.* 9, 494–500; Fig. 1.

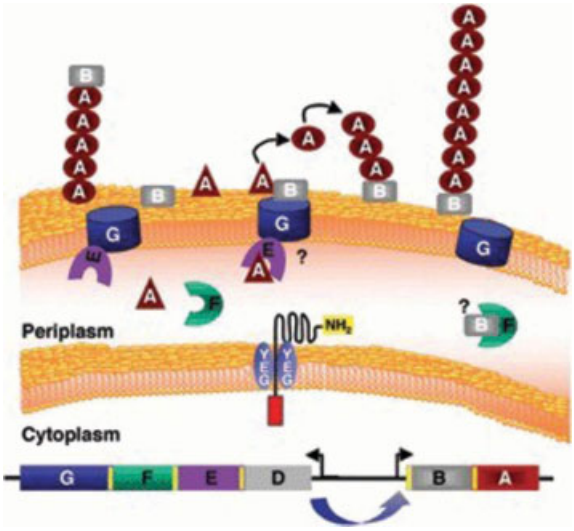


Fig. 8.10 Current model of curli fiber regulation and assembly. All curli subunits except CsgD start with a signal sequence required for translocation into the periplasm by the Sec pathway. CsgA and CsgB are the major and minor curli subunits, respectively, kept in a polymerization-competent form by the CsgF chaperone. CsgG might form a pore in the outer membrane through which CsgB and CsgA escape where CsgB acts as a nucleator. CsgE could transfer CsgA and CsgB to CsgG as indicated. M.R. Chapman, et al. **2003**, *ASM News* 69, 121–126; Fig. 2.

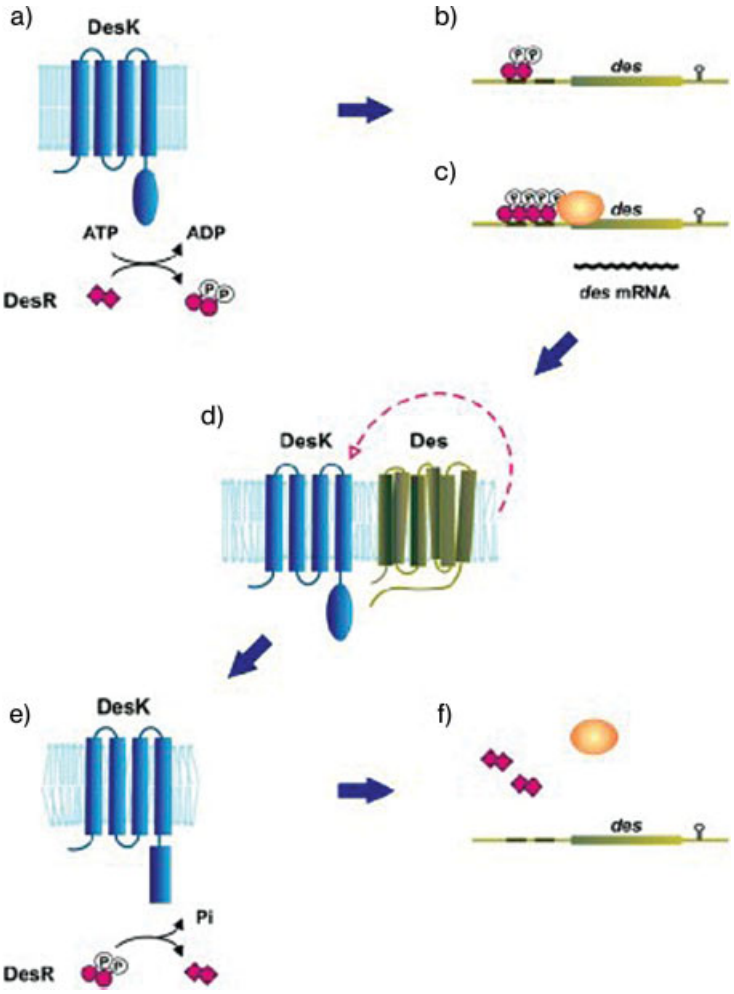


Fig. 9.6 Model of the Des pathway. (a) If cells experience a temperature decrease to 25 °C, this is sensed by DesK and leads to the activation of its kinase. DesK undergoes autophosphorylation and then transfers the phosphoryl group to the response regulator DesR. (b, c) The active DesR~P binds to DNA sites immediately upstream of the *des* promoter to initiate transcription by interaction with the RNA polymerase. (d) The desaturase inserts into the cytoplasmic membrane and introduces double bonds into acyl chains of the phospholipids which converts DesK from a kinase to a phosphatase. (e) Dephosphorylation of DesR~P results in turn-off of *des* transcription. M.C. Mansilla, et al. **2005**, *Arch. Microbiol.* 183, 229; Fig. 2.

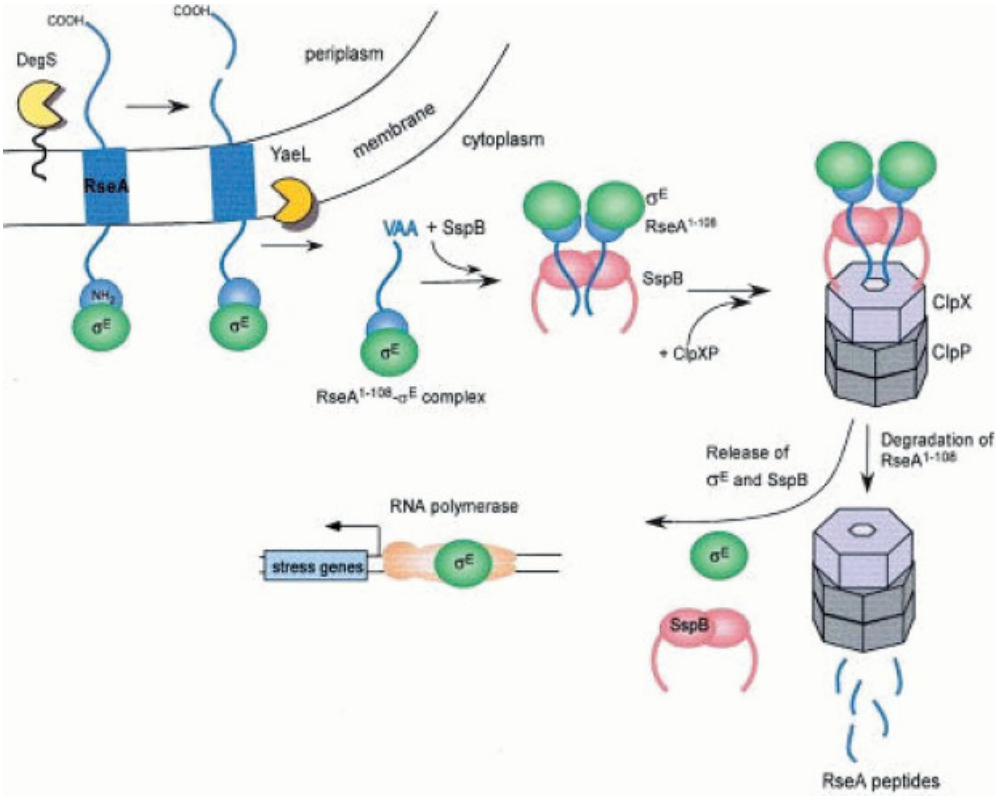


Fig. 9.20 Activation of σ^E needs the successive action of three different proteases. One way to release σ^E from its anti-sigma factor RseA involves the appearance of a subset of outer membrane proteins carrying the YQF signature at their immediate C terminus. In its unfolded form, this signature is recognized by the PDZ domain of the membrane-anchored DegS protease. This in turn leads to activation of the proteolytic activity of DegS, cleaving RseA in its periplasmic domain. Next, the RseP protease cuts within or near the transmembrane segment, causing release of the N-terminal part of RseA with σ^E still bound. The C terminus of the truncated RseA is recognized by the SspB adaptor protein targeting RseA to the ClpXP protease, where it is unfolded and degraded into peptides. The last proteolytic step finally leads to the release of σ^E. B.M. Alba, C.A. Gross **2004**, *Mol. Microbiol.* 52, 613–619. J.M. Flynn, et al. **2004**, *Genes Dev.* 18, 2292–2301; Fig. 4.

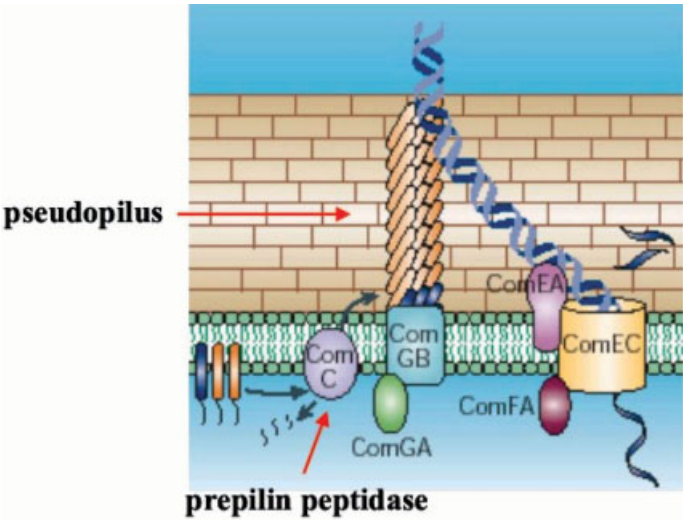


Fig. 10.1 The *B. subtilis* competence machine. The pseudopilus consists of the major pseudopilin (ComGC) and three minor pseudopilins (ComGD, ComGE, ComGG) which are processed by the prepilin peptidase (ComC). The polytopic membrane protein ComGB and the traffic NTPase ComGA assist this process. DNA molecules first make contact with the pseudopilin, then bind to the receptor protein, ComEA, which delivers the DNA to the ComEC channel located in the cytoplasmic membrane. Only one strand enters the cytoplasm; and the ATP-binding protein ComFA is involved in the transport process. I. Chen, D. Dubnau **2004**, *Nat. Rev. Microbiol.* 2, 241; Fig. 1, modified.

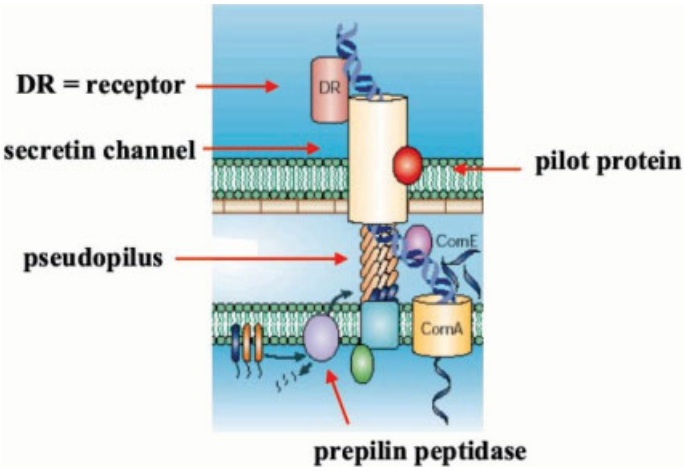


Fig. 10.2 The competence machine of *Neisseria gonorrhoeae*. This machine consists of a secretin forming a ring-like structure in the outer membrane which is connected to the pseudopilin (PilE, ComP), the subunits of which are processed by the prepilin peptidase (PilD). DNA molecules are recognized by the hypothetic receptor protein (DR) and delivered to the secretin (PilQ). Transport occurs through the water-filled secretin channel with the assistance of the protein PilP. The incoming DNA is recognized by a periplasmic DNA-binding protein (ComE) and delivered to a channel in the inner membrane (ComA). Only one strand enters the cytoplasm; and the other is degraded in the periplasm. I. Chen, D. Dubnau **2004**, *Nat. Rev. Microbiol.* 2, 241; Fig. 1, modified.

1
Structure of the Bacterial Cell

All bacterial cells are surrounded by at least one membrane, the cytoplasmic membrane enclosing the cytoplasm. This simple enclosure can be found only by species living within eukaryotic cells such as *Mycoplasma* spp. But most cells are surrounded in addition by a thick cell wall (the Gram-positives) and another group by a thin cell wall followed by a second membrane, the outer membrane (the Gram-negatives), where both membranes are separated by the periplasm. Furthermore, cells can be surrounded by an extracellular matrix, a capsule or an S-layer, a monomolecular protein layer. In total, Gram-positive bacteria consist of three compartments, the cytoplasm, the cytoplasmic membrane and the extracytoplasm, while Gram-negatives contain two additional compartments, the periplasm and the outer membrane. In addition, cells can contain appendages, flagella and/or pili allowing movements and attachment to cell surfaces, respectively. We will present an overview of the composition and function of these different compartments. For details, classic microbiology textbooks should be consulted.

1.1
The Cytoplasm Compartment

First of all, the cytoplasm contains all the different types of nucleic acids: the chromosome(s) and quite often one or more plasmids and the arsenal of RNAs (mRNA, rRNA, tRNA, sRNA). In addition, the cytoplasm consists of an aqueous solution of salts, sugars, amino acids, vitamins, coenzymes and a wide variety of other soluble materials including about 2000 different protein species. The cytoplasmic proteins fold into structures that place their hydrophobic segments in the interior to reduce their entropic cost of their exposure to water, and polar residues are generally placed on the protein surface where they can be hydrated. Traditional textbook representations of the prokaryotic cytoplasm show an amorphous amalgamation of proteins in which a randomly arranged chromosome is located. This image of the microbial cell is not correct and, instead, the cytoplasm is highly structured with many proteins carrying out their assigned functions at specific subcellular locations. Therefore, the bacterial cell is more than a bag of proteins. While most proteins may be dispersed throughout the cytoplasm, some seem to

Dynamics of the Bacterial Chromosome. Structure and Function. W. Schumann
Copyright © 2006 Wiley-VCH Verlag GmbH & Co. KGaA
ISBN 3-527-30496-7

have specific localizations. One example is the *B. subtilis* replisome which is present throughout replication of the chromosome within the middle of the cell. Proteins can be localized within the cell through a so-called localization tag which is fused either to the N- or to the C-terminus. The most prominent tag is GFP (for green fluorescent protein) and its variants. Furthermore, all macromolecular processes occur in the cytoplasm, such as DNA replication, transcription, translation and recombination.

The Bacterial Cytoskeleton

All eukaryotic cells, from yeast to plants to animals, have an internal framework called the cytoskeleton. The three types of strut that make up this framework – microtubules, actin filaments and intermediate filaments – not only provide cells with mechanical support, but also serve as tracks for motor molecules to move along. Bacteria were long thought to lack cytoskeletal filaments, suggesting that the cytoskeleton might have evolved after the first primitive eukaryotic cell developed from its bacterial origins. But work carried out over the past few years has shown that bacteria contain cytoskeletons with homologs to all the eukaryotic cytoskeletal elements. Bacteria contain a microtubule (tubulin) protein called FtsZ, multiple actin-like proteins including MreB and Mbl, and intermediate filament proteins such as crescentin in *C. crescentus*, indicating the very ancient origin of these types of intracellular filaments. Strikingly, no nucleators or motor proteins have been identified for any bacterial cytoskeletal element. In eukaryotes, microtubules are involved in various transport processes and, in forming the mitotic spindle during chromosome segregation, actin forms a dynamic cytoskeleton involved in cell shape regulation and mobility as well as the cytokinetic ring during cell division; and intermediate filaments provide mechanical support to the cell and nucleus. In procaryotes, the tubulin homolog FtsZ is essential for cell division, forming a cytokinetic ring at mid-cell early during the division process (see Section 3.3). In association with a number of other cell division proteins, FtsZ then constricts concomitantly with septal peptidoglycan synthesis to bisect the cell. The two actin-like proteins, MreB and Mbl, in *B. subtilis* form helical filaments. The two proteins do not interact to form a hybrid filament but each protein forms a filament of different pitch and length. While MreB forms a short (0.73 ± 0.12 μm) pitch filament that assembles around the mid-cell position, Mbl filaments (half-life ≈ 8 min) have a longer pitch (1.7 ± 0.28 μm) and traverse the entire cell length. MreB spirals are dynamic throughout the cell cycle, compacting at the division plane in predivisional cells, remaining there until division is complete and then expanding to fill the cell. MreB has a role in the maintenance of the cell width and plays a role in chromosome segregation, whereas Mbl is involved in the maintenance of the long axis of the cell. Depletion of either MreB or Mbl causes aberrant partitioning of the origins of replication and a large fraction of anucleate cells. The intermediate filament crescentin (encoded by *creS*) of *C. crescentus* is closely juxtaposed with the cell membrane and is responsible for the vibrioid shape: cell lacking crescentin are rod-shaped. CreS filaments represent true architectural

structures within the cell, as their conformation dictates cell shape. And the ParM protein encoded by plasmid R1 forms F-actin-like cables that actively move plasmid DNA. In 2001, the three-dimensional structure of MreB of *T. maritima* was solved and revealed that it is congruent with that of eukaryotic actins. Later, the structure of a plasmid-encoded ParM protein was solved and shown to be very similar to that of actin and MreB. Both proteins and several others, such as Mbl of *B. subtilis*, now belong to the MreB family; and these proteins carry out nucleotide binding and hydrolysis which regulates polymerization into linear protofilaments.

1.2 The Cytoplasmic Membrane Compartment

The cytoplasmic membrane is an 8 nm thick structure enclosing the cytoplasm, separating the interior of the cell from the environment and preventing the diffusion of substances into and out of the cytoplasm. It acts as a highly selective barrier to concentrate metabolic compounds and nutrients within the cytoplasm and to secrete waste products.

Structure and Composition of the Cytoplasmic Membrane

The cytoplasmic membrane consists of equal amounts of phospholipid and protein. It contains 65–75% of the cellular phospholipids and 6–9% of the cellular protein. The phospholipids consist of a hydrophobic part (fatty acids) and a hydrophilic part (glycerol), where the glycerol backbone contains two bound fatty acids and a phosphate group. Three major phospholipid species are present in *E. coli*, amounting to about 2×10^7 molecules per cell: 70–80% phosphatidylethanolamine, 15–25% phosphatidylglycerol and 5–10% cardiolipin. All of the phospholipids contain *sn*-glycerol-3-phosphate esterified with fatty acids at the *sn*-1 and *sn*-2 positions. The predominant fatty acids are the saturated palmitic acid (16:0) and the unsaturated species palmitoleic acid (16:1) and *cis*-vaccenic acid (18:1). The structure of the membrane is stabilized by hydrogen bridges and hydrophobic interactions; and, additionally, cations such as Mg^{2+} and Ca^{2+} help to stabilize the membrane by forming ionic bonds with the negative charges of the phospholipids. Fatty acid content (chain length, saturation) are dependent on the environmental conditions, including temperature (see Section 9.3), stage of growth and composition of the growth medium. In particular, membranes have to be maintained in a fluid, liquid–crystalline state to allow lateral diffusion of their proteins and protein complexes. In addition, fluid membranes have much higher permeabilities to small molecules than do gel-phase bilayers. The phospholipids arrange in a lipid bilayer, forming a hydrophobic barrier, preventing the uncontrolled movement of polar molecules and allowing the retension of metabolites and proteins.

While some proteins are tightly bound to the membrane with one or more membrane-spanning domains (so-called *integral membrane proteins*), others are loosely bound (peripheral membrane proteins) or interact only transiently. Bioin-

formatic analysis of the *E. coli* proteome indicates that approximately 1000 of the 4288 predicted genes encode integral membrane proteins that are critically important for many cellular functions. However, owing to their hydrophobic and amphiphilic nature, membrane proteins are difficult to study; and they account for less than 1% of the known high-resolution protein structures. Topology models describe the number of transmembrane spans and the orientation of the protein relative to the lipid bilayer. For *E. coli* proteins, location of their C-terminus can be easily determined through the use of the topology reporter proteins alkaline phosphatase (PhoA) and GFP. PhoA and GFP have opposite activity proliles: PhoA is active only in the periplasm, whereas GFP is fluorescent only in the cytoplasm. When fused to the C-terminus of a membrane protein, PhoA and GFP accurately report on which side of the membrane the C-terminus is located. The distribution of the membrane proteins changes considerably under different medium and growth conditions. The amino acid residues of membrane proteins in the interior have similar hydrophobic character, as residues in the interior of the soluble proteins and the residues that are exposed to the aqueous environment have the expected polar character. Residues exposed to the nonpolar lipid acyl chains have even greater aggregate hydrophobic character than residues in the protein interior important for maintaining the correct conformation and penetration of the protein in the nonpolar lipid bilayer. In general, transmembrane segments are considered to be α-helical. Cytoplasmic membrane proteins can be classified as to their function they exert:

- proteins involved in energy generation and conservation
- proteins involved in solute transport
- proteins involved in carbohydrate translocation
- proteins and protein complexes involved in the translocation of proteins through the cytoplasmic membrane
- other cytoplasmic membrane proteins.

Proteins Involved in Energy Generation and Conservation

Most biosynthetic and transport processes in eubacteria are driven neither by the hydrolysis of the high-energy phosphate bonds in ATP, GTP or phosphoenolpyruvate (PEP), nor by coupling to transmembrane ion gradients. Cells growing in fermentative conditions (absence of oxygen or other inorganic electron acceptors) produce ATP by substrate-level phosphorylation reactions in the glycolytic pathway. The ATP synthesized by these reactions can be used to form transmembrane ion gradients, primarily by reversal of the F_1F_0 proton-translocating ATPase. For cells grown in nonfermentive, respiratory conditions, the passage of electrons through an electron transfer chain to suitable electron acceptors (oxygen, fumarate, nitrate, nitrite, dimethyl sulfoxide [DMSO], trimethylamine *N*-oxide [TMAO] or hydrogen) is coupled to the extrusion of proteins and the creation of a transmembrane electrochemical gradient of protons. This proton motive force (PMF, $\Delta\mu_H$) can be used by many transport processes. Bacterial respiratory chains act as a series of physically separate protein complexes. Numerous membrane-bound

dehydrogenases transfer two electrons or hydrogen atoms from their specific substrates to the pool of quinones, which serve as mobile hydride carriers diffusing through the membrane. These quinones shuttle reducing equivalents from the dehydogenases to the terminal reductases or oxidases. While ubiquinone-8 is the predominant species in aerobic cells, menaquinone-8 is the major species in cells grown under anaerobic conditions.

PMF and ATP synthase are involved in energy conservation. The energy released by the passage of reducing equivalents along the electron transfer chains is captured in the form of an electrochemical gradient of proteins released upon quinol oxidation. PMF consists of two components, an *electrical potential* owing to the separation of charge and a *chemical gradient* or pH gradient, when the external pH differs from the internal pH. The magnitude of the electrical potential, $\Delta\Psi$, is measured from the distribution of permeant ions, while the pH gradient is measured from the distribution of permeant weak acids or bases. The multisubunit F_1F_0 proton-translocating ATPase, also called ATP synthase, carries out ATP synthesis of ATP at the expense of the transmembrane PMF generated in respiring cells. Alternatively, ATP synthase can use ATP generated by fermentative substrate-level phosphorylation to create the PMF. The enzyme is composed of eight subunits organized in two distinct complexes, the membrane-embedded F0 complex (subunits *a*, *b*, *c* in a 1:2:6–12 stoichiometry) forming a transmembrane proton channel and the peripherally bound F_1 complex. The F_1 complex consists of five subunits ($\alpha_3\beta_3\gamma_1\delta_1\varepsilon_1$).

Proteins and Protein Complexes Involved in the Translocation of Proteins Through the Cytoplasmic Membrane

Integral cytoplasmic proteins and proteins translocated through the cytoplasmic membrane need assistance by at least one protein. Two major pathways exist to translocate polypeptide chains through the cytoplasmic membrane, the Sec and the Tat pathways, accepting unfolded and folded proteins, respectively (see Section 8.4), while the insertion of proteins into the cytoplasmic membrane is assisted by YidC alone or in conjunction with the Sec pathway (see Section 8.4.3). In addition, in Gram-negative bacteria, there are protein complexes attached and integrated within the cytoplasmic membrane working in conjunction with outer membrane complexes destined to secrete proteins into the medium (see Section 8.6).

Proteins Involved in Solute Transport

Some small nonpolar and fat-soluble substances, such as fatty acids, alcohols and benzene, may enter and exit the cell readily by becoming dissolved in the lipid phase of the membrane. Charged molecules, in contrast, such as organic acids and inorganic salts, which are hydrophilic, do not readily pass the membrane barrier but instead must be specifically transported. Water freely penetrates the membrane, since water molecules are sufficiently small and uncharged to pass between phospholipids. Therefore, polar molecules have to be transported through

the action of membrane transport proteins. Such active transport mechanisms allow the accumulation of solutes against a concentration gradient. Most solutes are transported by an energy-dependent process and bacteria produce a remarkable number of active transport systems exhibiting high substrate affinity and specificity with a broad range of mechanisms of energy-coupling including:

- systems driven by *symport* or *antiport* with ion gradients
- *periplasmic permeases* as part of multiprotein complexes that consist of a periplasmic substrate-binding protein and are driven by ATP hydrolysis
- ion transport driven by *P-type ATPases*
- *serial transport systems* that mediate active transport across both the inner and the outer membrane
- *group translocation* processes that carry out the simultaneous transport and modification of their substrate.

Serial transport systems can mediate active transport across both the inner and the outer membrane. Since these systems are composed of components residing in three different compartments (the inner membrane, the outer membrane, the periplasm), they are explained under Section 1.5. Ion *symport systems* use a single integral membrane protein to couple substrate accumulation to the downhill movement of a driving ion in the same direction as the substrate. These electrogenic transport processes can be driven by proton or sodium gradients; and the lactose transporter, LacY, is the best studied ion-driven active transport system. The LacY permease contains 12 transmembrane segments of about 20 amino acids each, with both protein termini facing the cytoplasm. A stoichiometry of one proton per lactose molecule is seen under most conditions. Other symporters couple substrate accumulation to the sodium ion gradient which is established in enteric bacteria by the action of sodium/proton antiporters extruding Na^+ at the expense of PMF. One example is the MelB protein which transports melibiose and other α-galactosides coupled to Na^+ or H^+ gradients, while transport of methylthio-β-galactoside is coupled only to the Na^+ gradient. *Antiporters* couple the uptake of one compound to the release of a second type. Examples from *E. coli* are the potassium/proton exchanger Kha and the two sodium/proton exchangers NhaA and NhaB. Other antiport systems exchange the substrate for a metabolic process for its structurally similar product. *E. coli* cells growing anearobically with fumarate as electron acceptor synthesize a specific C4-dicarboxylate transport system that exchanges succinate for a molecule of the substrate, fumarate.

ATP-driven transport systems include periplasmic permeases, serial transport systems and P-type ATPases. Periplasmic permeases exhibiting high affinity for their substrates (K_m 0.1–1.0 μM) differ from symporters by virtue of their dependence on the high-energy phosphate pool. Their transport activity is relatively unaffected by protonophores dissipating the PMF. Periplasmic permeases, also called ABC (for *A*TP-*b*inding *c*assette) transporters appear to be made of four protein domains or subunits: two hydrophobic membrane-spanning domains (MSDs) that are presumed to constitute the translocation pathway or channel across the membrane and two hydrophilic nucleotide-binding domains (NBDs) that interact

at the cytoplasmic surface to supply the energy for active transport. ABC transporters now constitute one of the largest superfamilies of proteins known: There are 80 ABC transporters in the Gram-negative bacterium *E. coli*. In the transport classification database (http://www-biology.ucsd.edu/_msaier/transport/), the ABC family is currently subdivided into 22 subfamilies of prokaryotic importers and 24 subfamilies of prokaryotic exporters. All bacterial ABC transporters that mediate uptake utilize a high-affinity solute binding protein which is located in the periplasm of Gram-negative bacteria and is either tethered to the cell surface or fused to the transporter itself in Gram-positive bacteria. These binding protein-dependent transporters take up a wide variety of substrates, which include nutrients and osmoprotectants that range from small sugars, amino acids and small peptides to metals, anions, iron chelators (siderophores) and vitamin B_{12}. The best studied example is the maltose uptake system (Fig. 1.1). Maltose or maltodextrins will bind to the periplasmic maltose-binding protein (MBP, encoded by *malE*) present in its open form. This leads to a conformational change, the closed form, and enables interaction with the transporter $MalFGK_2$. Binding to the transporter in turn stimulates a conformational change back to the open form, opening the transporter at the periplasmic site with concomitent release of maltose into the transporter and binding ATP to the ATP-binding cassette, the MalK protein. Since the open form of MBP has a low affinity towards the transporter, it dissociates, thereby triggering closing of the transporter at the periplasmic site and opening at the cytoplasmic site driven by ATP hydrolysis.

ABC transporters also function in efflux of substances from bacteria, which include surface components of the bacterial cell (such as capsular polysaccharides, lipopolysaccharides, teichoic acid), proteins involved in bacterial pathogenesis (such as hemolysin, heme-binding protein, alkaline protease), peptide antibiotics, heme, drugs and siderophores. *E. coli* is thought to possess 57 ABC transporters, which can be phylogenetically divided into ten subfamilies. As 44 of these transporters have associated periplasmic-binding proteins, they are thought to be involved in solute transport.The other 13 transporters are not linked to periplasmic-binding proteins and may therefore be involved in export. CCmABC, which is involved in heme export, is the only one for which export function has been identified. The LolCDE complex is a unique lipoprotein-detaching apparatus, but not an exporter. It releases lipoproteins from one leaflet of lipid bilayers but does not transport them across membranes, representing a new ABC transporter class.

P-type ion-translocating ATPases consist of a large (ca. 100 kDa) protein subunit that is phosphorylated on an aspartate residue by ATP during the transport cycle and a smaller subunit associated with the catalytic subunit. Examples are multiple transport systems for potassium in *E. coli* (TrkD, TrkG, TrkH), three magnesium transport systems where MgtA and MgtBC mediate only Mg^{2+} influx while the CorA mediates bidirectional flux of Mg^{2+} and the uptake of Co^{2+}. A-type ATPases mediate resistance to toxic oxyanions, such as arsenate, arsenite and antimonite. These plasmid-encoded heavy metal resistance determinants are widespread in bacteria.

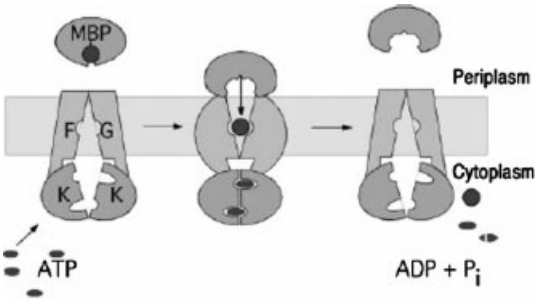


Fig. 1.1 The maltose uptake system. Maltose (represented by a black dot) is first bound by the periplasmic maltose-binding protein (MBP) forming a closed complex. Upon contact with the maltose transporter composed of the integral membrane proteins MalF and MalG, MBP opens to release maltose. Binding of MBP also triggers a conformational change in the ATP-binding cassette, leading to the binding of ATP. Next, maltose is transported into the cytoplasm, driven by ATP hydrolysis. A.L. Davidson, J. Chen **2004**, *Annu. Rev. Biochem.* 73, 241–268; Fig. 1.

Group translocation systems mediate simultaneous transport and modification of their substrates. The best characterized systems are PEP:sugar phosphotransferase systems (PTS), which use a cascade of phosphate transfer reactions from PEP via common and sugar-specific proteins to the sugar substrate during transmembrane uptake. Sugar accumulation occurs because the sugar–phosphate product is trapped inside the cell. Furthermore, the PTS plays a key role in catabolite repression and inducer exclusion coordinating and regulating carbohydrate metabolism.

Water fluxes across the cytoplasmic membrane are accomplished by two distinct mechanisms: (1) simple diffusion of water across the lipid bilayer and (2) diffusion through water-selective channels called aquaporins (AqpZ). AqpZ mediate rapid and large water fluxes in both directions in response to sudden osmotic up- and downshifts.

1.3 The Cell Wall Compartment

Most bacterial cells are surrounded by a rigid wall that has been thought to determine the shape of the cells (see Section 1.1). In both Gram-negative and -positive cells, the cell wall is located on the outside of the inner membrane, but is further surrounded by the outer membrane in Gram-negative bacteria. The major components of the bacterial cell wall are long glycan strands that are cross-linked by short peptides containing amino acids in both the D- and L-isoform and the whole ensemble is called *peptidoglycan* or *murein*, forming the *murein sacculus*. The murein of *E. coli* is composed of glycan strands with alternating *N*-acetylglucosamine and *N*-acetylmuramic acid sugar subunits linked together by β1→4 glycosidic bonds. Attached to the carboxyl group of each muramic acid by an amide linkage is a short peptide, L-alanyl-D-isoglutamyl-L-*meso*-diaminopimelyl-D-alanine, and

the whole unit is called muropeptide. The average glycan strand is bout 30 muropeptides in length. A molecule of lipoprotein (*lpp*) is attached to about every tenth muropeptide; and the covalent link is from the L-carboxyl group of DAP to the epsilon group of the carboxyl-terminal lysine of the lipoprotein. Further contacts occur with the porins OmpC and OmpF, providing together over 400000 possible contacts between the murein sacculus and the outer membrane. Elongation of the murein sacculus is a multisite process, which involves the insertion of strands at about 200 separate locations simultaneously; and the enzymes responsible move continously in one direction around the circumference. The precursors for peptidoglycan synthesis are a well conserved family of disaccharide pentapeptides called *lipid II*. They are synthesized in the cytosol and then flipped to the outer leaflet of the inner membrane, where they are added to the peptidoglycan meshwork by two distinct enzyme-catalyzed reactions. One of them, designated transglycosylation, adds disaccharide units to extend the linear glycan strand. In the other reaction, transpeptidation, interstrand cross-links form between juxtaposed peptide side-chains; and this reaction is the target of β-lactam antibiotics. Since the transpeptidases were first identified by their ability to bind penicillin, they are frequently called ''penicillin-binding proteins'' (PBPs). Most bacterial species code for multiple PBPs, where *E. coli* encodes about ten and *B. subtilis* about 16 PBPs. Peptidoglycan in *E. coli* is synthesized ubiquitously across the bacterial length, with the exception of the old poles, at which no new peptidoglycan synthesis occurs. Gram-positive bacteria contain another major polymer in addition to peptidoglycan in their cell walls, called teichoic acid. This anionic polymer is essential for Gram-positive bacterial cell wall viability, but its precise function is unknown. Bacterial cell walls provide protection against osmotic pressure in the cytoplasm and allow cells to withstand elevated hydrostatic turgor pressure.

1.4 The Outer Membrane Compartment

Gram-negative bacteria are surrounded by an additional membrane layer, the outer membrane whose most important function is to serve as a selective permeation barrier. It prevents the entry of noxious compounds and allow the influx of nutrient molecules. The outer membrane is an asymmetric lipid bilayer where the inner leaflet consists of phospholipids, while the outer leaflet is composed of lipopolysaccharide (LPS). As such, lipid bilayers show little permeability for hydrophilic solutes, including most nutrients. Amino acids, short peptides, sugars or oligosaccharides and other permeants below a threshold size of ~600 Da can cross the outer membrane by diffusion through porin channels. In contrast, other valuable or essential nutrients, including vitamin B_{12} and iron siderophore complexes, are too large and usually too scarce to be aquired effectively through porins. Substrate-specific, high-affinity active transporters are used for their uptake located in the outer membrane. The energy required to allow these transporters to deliver their substrates into the periplasm derives from a complex of proteins (TonB,

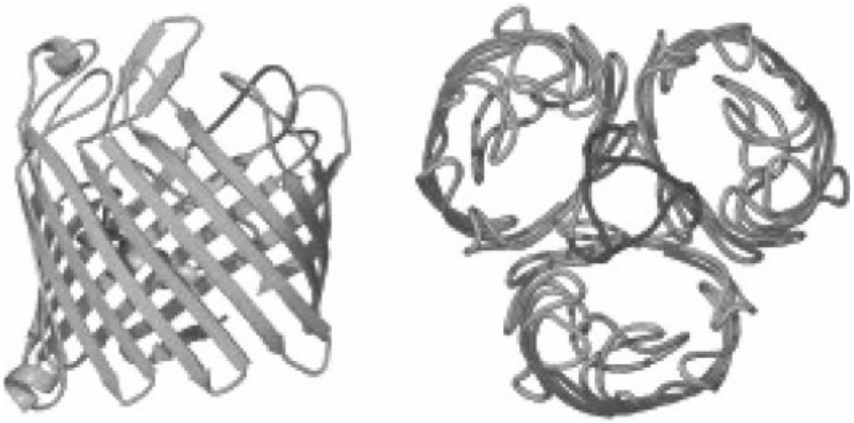


Fig. 1.2 Structure of the OmpF porin. View of the monomer from the side (left) and view of the trimer from the top. The loop inside the opening narrows the channel. H. Nikaido **2003**, *Microbiol. Mol. Biol. Rev.* 67, 593–556; Fig. 2. (This figure also appears with the color plates.)

ExbB, ExbD) in the cytoplasmic membrane (see Section 1.5). Infective agents such as some colicins and phages evolved to parasitize TonB-dependent systems to kill target bacteria.

The outer membrane of Gram-negative bacteria such as *E. coli* contains three major classes of proteins: lipoproteins, β-barrel proteins and multicomponent surface structures, such as pili and flagella. Liporoteins have lipid covalently attached to the N-terminal cysteine that anchors these molecules in the inner leaflet of the outer membrane. The β-barrel proteins are composed of β-sheets that are wrapped into cylinders (Fig. 1.2). Because of this structure, many of these proteins, also called porins, form channels allowing the influx of nutrients and the extrusion of waste products. These porins can be classified in two groups: nonspecific and specific channel-forming proteins. Nonspecific porins such as OmpF act as molecular sieves, while the others, like LamB, exhibit substrate specificity, in this case for maltose polymers. Since the β-barrel proteins are exposed on the surface, they may be used as phage receptors, such as OmpF for K20 and LamB for λ. The third major class of outer membrane proteins, the pili, serve to attach bacteria to surfaces. Each of the many different types of pili has specific affinity for certain surfaces (see below).

Composition and Biological Function of Lipopolysaccharide

Lipopolysaccharide (LPS) is a unique constituent of the bacterial outer membrane and is composed of three components: the proximal, hydrophobic lipid A region, the distal, hydrophilc O-antigen polysaccharide region that protrudes into the medium, and the core oligosaccharide region that connects the other two. Lipid A is a polar lipid of unusual structure, in which a backbone of glucosaminyl-β-(1→6)-glucosamine is substituted with six or seven fatty acid residues, all of them

saturated. LPS biosynthesis begins in the bacterial cytoplasm as the acylation of uridyl-diphospho-*N*-acetylglucosamine (UDP-GlcNAc) with β-hydroxymyristate. After deacetylation, the product of this reaction is further modified with β-hydroxymyristate to generate UDP-2,3-diacylglucosamine. Cleavage of the pyrophosphate bond produces 2,3-diacylglucosamine-1 phosphate and, after condensation of this compound with another molecule of UDP-2,3-diacylglucosamine and 4′ phosphorylation, the intermediate lipid IVA is formed. Two Kdo (3-deoxy-D-manno-oct-2-ulosonic acid) residues are then transferred to lipid IV_A. Two acetyltransferases, HtrB and MsbB, add lauroyl and myristoyl groups. The core sugar residues are added on this intermediate and the MsbA ABC transporter is thought to translocate these molecules from the cytoplasmic side to the periplasmic surface of the plasma membrane. The O-antigen is added after the MsbA-mediated translocation. Subsequent transport reactions move the LPS molecules across the periplasmic space into the inner leaflet and finally to the outer leaflet. Lipid A is the biologically active component of LPS which causes inflammation and septic shock through Toll-like receptor 4 recognition.

Three kinds of LPS modifications have been observed: (1) substitution of the phosphate group at the 4′ position with phosphoethanolamine providing resistance against polymyxin B, a lipid A-binding cationic cyclic peptide antibiotic, (2) L-Ara4N decoration which is induced by activation of PrmA forming a two-component system (see Section 6.2.9) with its cognate histidine kinase, PrmA, and (3) addition of palmitate via acyloxyacyl linkage at position 2. LPS palmitoylation provides resistance against cationic antimicrobial peptides induced by the innate immune response to bacterial infections. These modifications may occur alone or in combination on a single LPS molecule, yielding at least six different LPS species.

Gram-negative bacteria have long been known to produce membrane blebs or vesicles of 0.5–1.0 μm into the culture medium when bacteria where grown in lysine- or phosphate-limited conditions or when protein synthesis was inhibited. The membrane component of these vesicles is derived from the outer membrane. The naturally competent *Neisseria* and *Haemophilus* spp produce large amounts of vesicles in culture to transfer DNA. Commensal *Bacteroides* spp of rumen intestine produce vesicles containing cellulase and xylanase which can provide the cells with a source of carbon from the digestion of nonmetabolizable carbohydrate polymers. Very recently, it has been shown that bacterial vesicles have an important biological function in the secretion and delivery of bacterial protein toxins to mammalian cells (see Section 8.6.6).

Outer Membrane Proteins

Outer membrane proteins (OMPs) are synthesized in the bacterial cytoplasm as precursors with N-terminal signal peptides and then translocated across the cytoplasmic membrane via the Sec pathway (see Section 8.4.1). After removal of the signal peptide by the signal peptidase, the mature proteins assume β-barrel structures with hydrophobic outer surfaces, suitable for interaction with LPS or other membrane lipids. Several factors are involved in the folding of OMPs, where most

of these are only in part required for the targeting of OMPs and presumably fulfill overlapping or partially redundant functions. Outer membrane proteins can be classified into several classes depending on their function:

- lipoproteins
- general porins
- substrate-specific porins
- other outer membrane proteins.

Lipoproteins

More than a dozen lipoproteins have been described where the murein lipoprotein (gene *lpp*) is the most prominent and best studied member. Lpp is a small protein (7200 Da) present in approximately 7×10^5 copies per cell. Its N-terminal cysteine is modified at two sites. Its sulfhydryl group is substituted with a diglyceride and its amino group by a fatty acyl residue, allowing penetration into the inner leaflet of the outer membrane. About one-third of the murein lipopotein molecules are bound covalently to the peptidoglycan layer through the ε-amino group of its C-terminal lysine, thereby fixing the outer membrane to the murein sacculus. Deletion or mutations of *lpp* result in numerous defects such as periplasmic leakage, increased susceptibility to many toxic compounds and the formation of membrane vesicles.

General Porins

The general diffusion pores formed by porins allow the diffusion of hydrophilic molecules (<600 Da) and exhibit no particular substrate specificity, despite some selectivity for either anions or cations. They form water-filled channels across the outer membrane and are either homotrimers formed by three hollow β-barrels, such as OmpC, OmpF and PhoE, or monomeric, including OmpA and OmpG. The classic porins OmpC, OmpF and PhoE have 18 antiparallel β-strands each. The OmpC and OmpF porins are among the most abundant outer membrane proteins and their expression is extensively regulated. OmpC, with a pore diameter of 1.08 nm, is thought to be important in environments where nutrients and toxin concentrations are high, such as in the intestine, and it is the predominant porin at high temperatures and high osmolarities. The three-dimensional structure of OmpF has been determined and the most conspicuous structural feature is the presence of an 'eyelet' region, a narrow site constricting the pore and lined with charged residues. These residues are expected to cause a strong transversal electric field, in addition to a screw-like field in the wider parts of the pore. This feature is conserved in general diffusion porins.

OmpF, which has a larger pore diameter (1.12 nm) which results in a 10-fold faster diffusion rate, is thought to be important in habitats where nutrient and toxin concentrations are low, such as in fresh water, and it is more abundant at low temperatures and low osmolarities. Both OmpF and OmpC play an important role in maintaining the periplasm and cytoplasm in an iso-osmotic state. Both proteins

are present at 10^5 molecules per cell. While in low-osmolarity medium OmpF predominates, OmpC is present at enhanced levels in high-osmolarity medium at the expense of OmpF; and the relative amount of each protein in response to the actual medium osmolarity is regulated by the EnvZ–OmpR two-component signal transduction system (see Section 6.2.8.3).

The trimeric phosphoporin PhoE is produced only under conditions of phosphate starvation. The channel-forming motif of PhoE is a 16-strand anti-parallel β-barrel. Short β-hairpin turns define the periplasmic end of the barrel, whereas long irregular loops are found at the cell surface.

OmpA (35 kDa) inserts as a monomer into the outer membrane and is present in about 100000 copies per cell. OmpA contains a nonspecific diffusion channel with a pore diameter of 0.7 nm; and the penetration of solutes through the OmpA channel is about two orders of magnitude slower than that through the OmpF channel. OmpA is composed of two domains, an N-terminal membrane-embedded domain of 170 amino acid residues, serving as a membrane anchor, and a C-terminal 155-residue domain located in the periplasmic space and proposed to specifically interact with the peptidoglycan layer. The N-terminal domain consists of eight antiparallel β-strands that are connected by three short periplasmic turns and four relatively long surface-exposed loops forming a hydrophobic protein surface and a polar interior, where the barrel interior accomodates several small water-filled cavities. But no continous transmembrane channel could be detected questioning the relevance of OmpA in pore formation. Mutants lacking OmpA are extremely poor recipients in conjugation experiments and they tend to produce spherical cells.

The monomeric OmpG is able to rescue the growth of porin-deficient bacteria on media containing maltodextrins as large as maltopentose as the sole carbon source. The physiological role of OmpG is not clear. OmpG is not detected in the outer membrane of several *E. coli* strains, such as K-12, by Western blot analysis. But the *ompG* gene is detected by PCR analysis in the genome of these strains, indicating that the lack of expression may be due to the growth conditions.

Substrate-specific Porins

Besides general diffusion pores that discriminate between solutes (if at all) on their charge and size, there is a second class of porins which recognize their substrate. Among them are LamB of *E. coli* and ScrY of *Salmonella typhimurium*. The *lamB* gene is part of the *mal* regulon (see Section 6.2.8.2) which is induced by maltose or maltodextrins. The LamB protein forms homotrimers whose monomers consist of 18-stranded antiparallel β-barrels. The substrate translocation pathway involves a row of aromatic amino acids (greasy slide) that is lined up by polar residues (ionic track). Sugar residues (maltose or maltodextrins up to maltoheptaose) are in van der Waals' contact with the greasy slide by their hydrophobic face, while hydrogen bonds are formed between their hydroxyl groups and the ionic track residues. Movement of the sugar through the channel proceeds by continous disruption and formation of these hydrogen bonds. In a chemostat study under car-

bohydrate-limiting conditions, the LamB production became strongly derepressed and conferred growth advantages at limiting concentrations of not only glucose but also lactose, arabinose and even glycerol (in addition to maltose, trehalose, melibiose). In summary, the LamB channel is not really a maltose-specific channel, but it facilitates the influx of a wide variety of carbohydrates when they exist in low concentrations in the environment. LamB serves also as the receptor for bacteriophage λ. Other specific channels are ScrY (allows the rapid diffusion of a large varieties of sugars, such as glucose, fructose, arabinose, maltose, lactose, raffinose, sucrose, maltodextrins), BglH (aryl-β-D-glucosides), Tsx (nucleoside transport) and FadL (utilization of long-chain fatty acids as a carbon source).

Other Outer Membrane Proteins

Since the outer membrane acts as a major barrier when proteins have to be exported, we need specific mechanisms to overcome this barrier. Currently, at least five different mechanisms are known to be utilized and these are described in detail under Section 8.6. Export channels of the type I machinery make use of the outer membrane channel TolC, while those of type II are called secretins. Components of P and type I pili are secreted across the outer membrane, using an usher which forms ring-like structures, 15 nm across, and contains a central pore.

OmpT of *E. coli* is a surface membrane protease and is the prototypical member of the omptin family of Gram-negative bacteria. OmpT catalyzes the activation of plasminogen to plasmin, a function that is physiologically relevant for the virulence of *Y. pestis* and for clinical *E. coli* isolates. OmpT also plays a role in bacterial virulence by the cleavage of protamine and other cationic peptides with antibiotic activity. The protease folds into a 10-strand antiparallel β-barrel conformation with extracellular loops that extend well beyond the membrane. The active site is located within a deep groove formed by loops L4 and L5 on the one side and L1, L2, and L3 on the other. The structure also revealed a binding site for a single LPS molecule that appears to be important for the catalytic activity of the enzyme. OmpT seems to function through a novel mechanism involving an Asp210–His212 catalytic dyad that, together with Asp83–Asp85, activates a putative nucleophilic water molecule. OmpT cleaves substrates between dibasic residues with high catalytic efficiency, which has been shown to be important for the inactivation of antibiotic peptides and colicins and the degradation of some recombinant proteins expressed in *E. coli*. It copurifies with protein inclusion bodies and retains activity under denaturating conditions, including boiling or in the presence of up to 4 M urea; and therefore it can be a major source of protein degradation during the solubilization and renaturation of inclusion bodies.

What is known about Omp biogenesis? Omps fold in the periplasm before their insertion into the outer membrane. In vitro, LPS stimulates the folding of Omps; and the periplasmic chaperone SurA also stimulates Omp folding, while another periplasmic protein, Skp, plays an as yet unidentified role. The insertion of proteins into membranes generally requires a proteinaceous machinery, but no components of such a putative machinery have been identified. A protein possibly ful-

filling these criteria is the surface antigen designated Omp85 in *Neisseria* spp and D15 in *Haemophilus* spp. Genes coding for Omp85/D15 homologs are present in all Gram-negative bacteria examined. The *omp85* gene is essential and, upon Omp85 depletion, unassembled forms of various outer membrane proteins accumulate within the periplasm.

What is known about the dynamics of the outer membrane surface in live cells? To answer this question, the spatial distribution of LamB found mainly as a trimer in the outer membrane (there are about 30000 monomeric copies) has been studied using fluorescently labeled phage λ tails. It turned out that there are two populations, mobile (~40%) and immobile LamB trimers. The mobile LamB trimers move along the bacterial surface and these movements are restricted by an underlying dynamic spiral pattern. The distribution of the protein changes within tens of seconds.

R. Koebnik, et al. **2000** Structure and function of bacterial outer membrane proteins: barrels in a nutshell. *Mol. Microbiol.* 37, 239–253.

1.5
The Periplasmic Compartment

The periplasmic space extends between the inner and outer membranes of Gram-negative bacteria. The architecture of this compartment facilitates cell wall and outer membrane growth and coordinates these processes with cell division. Proteins residing in the periplasm fulfil important functions in the detection and processing of essential nutrients and their transport into the cytoplasm. Furthermore, periplasmic proteins promote the biogenesis of proteins entering this compartment along with components destined for incorporation into the peptidoglycan, outer membrane or capsules. In addition, the periplasm can contain glucans which play an important role in adaptation to hypoosmotic conditions.

The width of the periplasmic space has been determined to vary over 13–25 nm, depending on the growth conditions. Located within this region is the peptidoglycan layer, with an estimated thickness of 5–8 nm, which is in close association with the outer membrane. Attachment of these two structures is facilitated by major outer membrane lipoprotein, Lpp, and by strong interaction of the porins with the peptidoglycan.

Periplasmic Glucans

Periplasmic glucans, also called MDOs (*m*embrane-*d*erived *o*ligonucleotides), are important components of the bacterial envelope of many Proteobacteria. Glucose is the sole sugar component and residues are linked essentially by β-glycosidic bonds. Glucans of *E. coli* are heterogeneous in size, varying from five to 12 glucose residues, with the principal species containing eight to nine glucose residues. Their structures are highly branched, the backbone consisting of β-1,2 linked glu-

cose units to which the branches are attached by β-1,6 linkages. These polymers are substituted to various degrees by *sn*-1 phosphoglycerol, phosphoethanolamine, phosphoethanolamine and succinic acid O-ester. In *E. coli*, the *opgGH* operon encodes two proteins (OpgG, OpgH) required for the assembly of the glucan structure. OpgH, a glycosyl-transferase anchored in the cytoplasmic membrane catalyzes linear oligomers of β-1,2-linked glucose units synthesized from UDP-glucose. OpgH could form a channel for glucan translocation from the cytoplasm to the periplasm during synthesis. OpgG is a periplasmic protein whose function might be to catalyze the addition of branches to the linear backbone emerging on the periplasmic side of the inner membrane. The two inner membrane proteins OpgB and OpgC have been implicated in glucan substitutions. While OpgB is a phosphoglycerol transferase with its activity located at the periplasmic face of the cytoplasmic membrane, OpgC is required for the periplasmic succinylation of the glucans. The *opgGH* operon is under osmotic control and many Gram-negative bacteria respond to hypoosmotic conditions by synthesizing large amounts of periplasmic glucans. Under conditions of low osmolarity they can account for up to 5% of the dry cell weight. The nature of the osmosensor is currently unknown.

Periplasmic Proteins

Periplasmic proteins can be divided into several categories based on their functions:

- solute or ion-binding proteins that function together with ABC transporters or chemotaxis receptors for the sensing and uptake of amino acids, vitamins, sugars, peptides and ions;
- catabolic enzymes that degrade complex molecules into smaller ones for transport through the inner membrane;
- detoxifying enzymes protecting cells from toxic compounds;
- enzymes involved in the biogenesis of envelope proteins and proteinaceous appendages, LPS, peptidoglycan, capsules and MDOs;
- serial transport systems.

Periplasmic binding proteins concentrate the solutes due to their high affinity for their ligand (in the range 0.1–1.0 μM) and their high concentration (up to 1.0 μM). Binding proteins are monomers generally consisting of two globular domains with a molecular mass of 20–40 kDa. Crystal structures of several binding proteins revealed that they are ellipsoidal and consist of two lobes connected by a flexible linker. These lobes are apart in the unliganded state and they come together to bind the solute, using hydrogen bonds. The complex binding protein–ligand interacts with the cognate membrane permease to promote the release of solute and its transport across the inner membrane. Four binding proteins play a role in chemotaxis (see Section 6.2.9): MalE interacts with Tar (specific for serine), MglB and RbsB with Trg (ribose and galactose) and DppA with Tap (peptides).

Polypeptides emerging on the outer surface of the inner membrane must fold into their native state and assemble into quaternary structures in the case of oligomeric proteins. Proteins destined for the outer membrane may be either released

as periplasmic intermediates before insertion or may be translocated directly from the inner to the outer membrane. Correct folding of periplasmic and outer membrane proteins is aided by protein disulfide isomerases catalyzing disulfide bond formation and proline isomerases (see Section 7.3). Some of these proteins may also act as molecular chaperones.

Serial Transport Systems

While most nutrients gain access to their transporters in the cytoplasmic membrane by diffusion through porins across the outer membrane (see above), a few substrates are too large to effectively enter by this route. Examples are the ferric siderophore complexes and vitamin B_{12}. Two serial transport systems, also called gated channels, have been described as being involved in the uptake of these compounds: TonB/ExbB/ExbD and Tol/Pal, which energize the transport via the PMF. Gram-negative bacteria contain several high-molecular-weight outer membrane proteins, so-called TonB-dependent receptors. Examples of such receptors include BtuB for vitamin B_{12} and six receptors for Fe^{3+}-siderophore complexes, including FhuA, FecA and FepA. They are involved in the uptake of large substrates, including iron-siderophore complexes (iron chelators of microbial origin) and vitamin B_{12}. These receptors bind to the ligands with high affinity and their function requires an interaction with the periplasmic protein TonB, which either spans the thickness of the periplasm or shuttles between the inner and outer membrane. The action of TonB requires that the cytoplasmic membrane is energized and that the energy is transferred to the receptors with the assistance of two cytoplasmic membrane proteins, ExbB and ExbD. The cellular ratio of the three proteins TonB:ExbB:ExbD is 1:2:7; and this stoichiomeric information predicts a complex of ~260 kDa.

How do these receptor proteins work? When the crystal structures of several siderophore receptors and BtuB were solved, the β-barrel in all these monomeric proteins consists of 22 strands and the N-terminal portion, consisting of about 150–200 residues, was found as a globular domain that is inserted into the barrel from the periplasmic side, forming a plug. Binding of a ligand induces a conformational change within the plug, so that the most N-terminal portion containing a short motif, the so-called TonB box (interacts with TonB) can interact with TonB. This first step does not create a large channel, but is followed by a large-scale conformational change caused by the energized TonB.

Two major models have been proposed for how TonB might act as a energy transducer: the propeller and the shuttle models. In the propeller model, TonB remains associated with the complex at all times and undergoes a rotary motion of the C-terminal propeller, initiated by ExbB, ExbD and the PMF. When the propeller becomes associated with the barrel of the outer membrane transporter, the rotary motion initiated at the inner membrane causes the outer membrane transporter to release its bound ligand into the periplasmic space. In the shuttle model (Fig. 1.3), TonB starts in an unenergized conformation in complex with ExbB and ExbD. ExbBD use the PMF to convert TonB to what is believed to be an energized

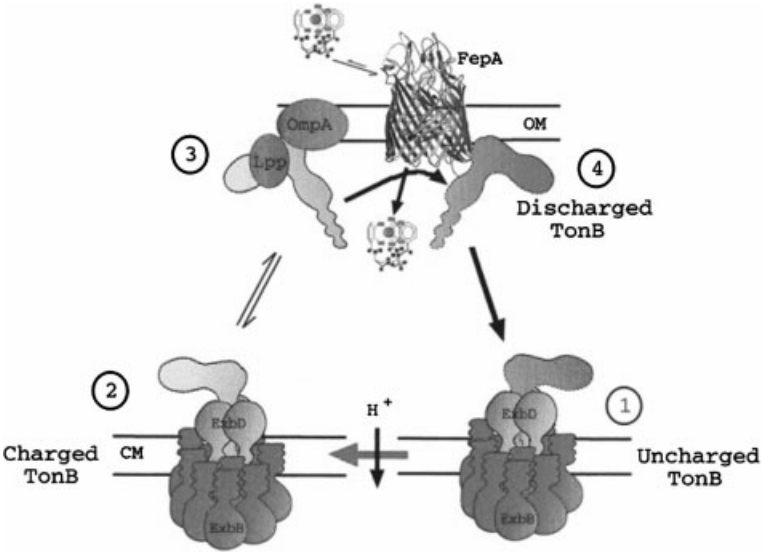


Fig. 1.3 Shuttle model for TonB-dependent energy transduction. Uncharged TonB (1) is energized by uptake of a proton (2) and shuttles to the outer membrane where is docks to OmpA/Lpp waiting for a transporter such as FepA with a bound ligand. Upon interaction with the Ton box of the the transporter the conformational energy is transduced to the transporter (3) triggering uptake of the ligand into the periplasm (4). The discharged TonB shuttles back to the ExbBD complex to become recharged. K. Postle, R.J. Kadner **2003**, *Mol. Microbiol.* 49, 869–882; Fig. 2. (This figure also appears with the color plates.)

conformation. Then, TonB diffuses to the outer membrane where it docks to either OmpA or Lpp waiting for transporters to signal ligand occupancy. Amino acid 160 of TonB interacts with the Ton box of the outer membrane transporter. Subsequently, TonB is recharged.

The Tol–Pal system forms a *trans*-envelope bridge linking inner and outer membranes and the peptidoglycan layer and is composed of seven proteins (Fig. 1.4). TolQ, TolR and TolA are inner membrane proteins that interact together by their transmembrane helices. The periplasmic TolB and the lipoprotein Pal form another complex anchored to the outer membrane. TolB interacts in a Pal-dependent manner with OmpA and Lpp, while Pal forms a homodimer in the cell envelope and interacts with OmpA. Pal is a lipoprotein with a serine residue at position +2 that results in its localization in the outer membrane. After cleavage of the signal sequence by the signal peptidase II and acylation, Pal is targeted to the outer membrane by the LolABCDE system (see Section 8.5). The peptidoglycan-binding sequence of Pal is located between residues 97 and 114 and forms a proposed α-helical motif.

Two functions have been suggested for the Tol–Pal apparatus. First, this system may contribute to maintaining cell envelope integrity through the interactions between the TolB and Pal proteins with the murein layer and the OmpA and Lpp proteins. Second, the Tol–Pal system might play a more dynamic function in cell

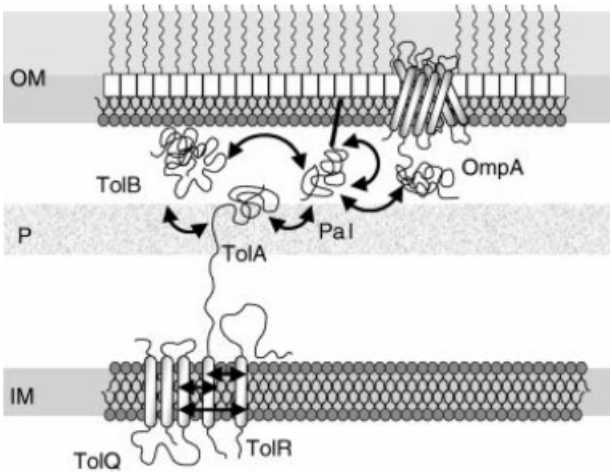


Fig. 1.4 The Tol–Pal serial transport system. The topologies of the different protein components are shown and interactions between these components indicated by arrows. E. Cascales, R. Lloubes **2004**, *Mol. Microbiol.* 51, 873–888; Fig. 9.

envelope biogenesis. It might function to bring the inner and outer membranes in close proximity via the TolA C–terminal domain–Pal interaction. The central domain of TolA, which forms an extended amphipatic α-helix, could be required for the transport of outer membrane components through the aqueous periplasm to reach the inner leaflet of the outer membrane. But it remains to be analyzed whether the PMF-dependent energization of the Tol–Pal system is directly involved in maintaining the cell wall architecture.

1.6 Extracellular Matrices

Surface Layers

Surface layers (S-layers) have been found in up to 400 different species of both eubacteria and archaea as the outmost cell envelope layer. They consist of one species of (glyco)protein, the S-protein, which assembles into characteristic two-dimensional crystalline layers (lattices) at the cell surface. This assembly is an entropy-driven process during which individual S-protein monomers form multiple interactions with each other and with the underlying cell envelope. The lattices can be quite porous, with pores occupying up to ~70% of their surface. The pores within an S-layer are of identical size (usually in the 2–8 nm range) and shape; and two or more distinct classes of pores can be present. S-layer proteins are among the most abundant cellular proteins and the S-layers fulfill many roles for the cell. They function as protective coats, as molecular sieves, as structures involved in cell adhesion and surface recognition, as scaffolding for enzymes, as

virulence factors and as molecular and ion traps. Furthermore, they act as virulence factors in pathogens, mediate resistance against bactericidal complement activity and adhesion to extracellular matrix proteins and temper the proinflammatory cytokine response. In summary, S-layers presumably evolved by selection in response to specific environmental and ecological pressures.

The molecular mass of S-layer proteins ranges from 40 kDa to 170 kDa, and some of them can become glycosylated (see Section 6.5.3). These proteins are often weakly acidic, typically contain 40–60% hydrophobic amino acids and possess few or no sulfur-containing amino acids. The S-proteins from Gram-negative bacteria interact with the sugar moiety of the LPS component of the outer membrane. It has been determined that S-proteins constitute between 5–10% of the total protein content of the cell, and $\sim 5 \times 10^5$ molecules are required to surround an average-sized bacterial cell. Most S-layer genes are monocistronic units expressed from one or more strong promoters. In most cases, the mRNA contains an untranslated 5′ region (ranging over 33–358 bp) which can fold into stable secondary structures causing a half-life of the S-layer mRNA from 10 min to 22 min. The biosynthesis of S-layers is a very complex process in which the amount of the protein component, its translocation through the cell wall and its incorporation into the existing S-layer lattice have to be coordinated with the growth rate of the bacterium and the synthesis of other cell wall components. To keep the surface of a bacterium with an average generation time of 20 min completely covered with the closed protein lattice, approximately 500 S-layer subunits have to be synthesized per second. Some eubacterial species can produce two superimposed S-layers; and each is usually composed of a different subunit species.

U.B. Sleytr, T.J. Beveridge **1999**, Bacterial S-layers, *Trends Microbiol.* 7, 253.

Capsules

Many eubacteria secrete on their surfaces slimy materials which in most cases consist of polysaccharides and, in a few cases, protein. These layers are called capsules and vary in different microorganisms; but they usually contains glycoproteins and a large number of different polysaccharides, including polyalcohols and amino sugars. Capsule layers have several functions:

- They play an important role in the attachment of certain pathogenic organisms to their hosts.
- They provide protection against phagocytic cells and the immune system. The most prominent case is that of *S. pneumoniae* where capsulated cells injected into a mouse will cause death in a few days.
- Since they bind a significant amount of water, they play some role in resistance to desiccation.

Capsular polysaccharides are high-molecular-weight acidic polymers composed of oligosaccharide repeating subunits and are found on bacterial cell surfaces and classified into four groups. In *E. coli*, the group-1 K capsular serotype of each

strain is dependent upon its repeat unit structure. Approximately 80 different K antigens have been identified in *E. coli*, reflecting differences in the structure and immunochemistry of the capsular polysaccharide repeat units. Here, the *E. coli* K30 serves as the prototype. K-antigen biosynthesis involves the production of undecaprenol pyrophosphate-linked repeat units by glycosyltransferases acting at the cytoplasmic face of the inner membrane. The repeats units are thought to be exported across the inner membrane involving a putative flippase (Wxz), where they are polymerized by a reaction requiring the *wzy* gene product. The majority of the undecaprenol pyrophosphate-linked repeat units are used for polymerization of capsular polysaccharide and require a tyrosine autokinase, Wzc, and its cognate phosphatase, Wzb. After polymerization, the capsule is transported to the cell surface in a process that involves the outer membrane lipoprotein Wza. This lipoprotein forms multimeric structures resembling secretions for type II and type III protein secretion (see Sections 8.6.2, 8.6.3) acting as channels in the outer membrane to allow the polymers to reach the cell surface. It is not yet clear how the polymer is moved through the periplasm to these channels, nor how it is assembled on the cell surface to form the capsular structure. Group 1 capsules afford protection from the host immune system by impeding phagocytosis. They are also involved in an intimate association with the glycocalyx of epithelial cells and probably contribute to biofilm development, but initial attachment to the epithelium also requires the participation of pili.

1.7 Appendages

Various types of exocellular appendages are important accessory components for bacteria to best fit their particular ecological niche. These appendages consist of flagella and pili (also called fimbriae). While flagella are involved in bacterial motility, pili mediate adhesion to specific targets in the environment. Both flagella and pili are anchored through the cell body to both periplasmic and membrane components.

Flagella

Flagella (singular: flagellum) are long, thin (about 20 nm diam.) appendages free at one end and attached to the cell at the other. They consist of the three main components basal body, hook and filament and function in the movement and chemotaxis of bacterial cells (Fig. 1.5). While some bacterial species contain one single, polar flagellum, others contain multiple flagella at one pole and others many flagella around the whole cell (peritrichous). Flagella are able to rotate exerted by a proton- or sodium-driven rotary motor that switches between counterclockwise and clockwise rotation.

The bacterial flagellum is composed of over 20 protein species with approximately another 30 proteins required for regulation and assembly; and it represents the most complex of all prokaryotic organelles (Fig. 1.5). The filament is ty-

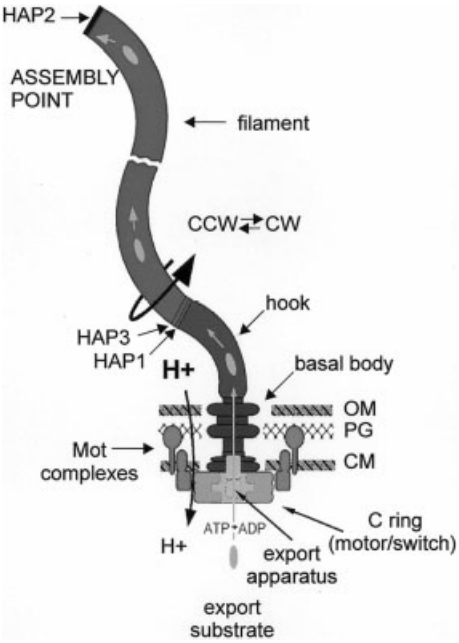


Fig. 1.5 Composition of the bacterial flagellum. The flagellum consists of the three components basal body (anchors the organelle within the cell envelope and functions as the switch), the hook and the filament, acting as a propeller. S.L. Bardy, et al. **2003**, *Microbiology* 149, 295–304; Fig. 1.

pically about 20 nm in diameter and usually consists of thousands of copies of a single protein called flagellin. In some cases, the filament is composed of several different flagellins. At the tip of the flagellum is a capping protein termed HAP2, forming a pentamer. New flagellin subunits assemble at the end of the filament under the cap protein, forming a plate-like structure with five legs protruding downwards and interacting with the filament. As each subunit is added, the cap rotates along the end of the helical filament, with a complete rotation of the cap occuring for every 55 flagellin subunits added. The filament is connected to the basal body by the hook, composed of a single protein, FlgE. The hook has a defined length of 55 nm and consists of about 120 subunits of FlgE. Hook length is controlled by the C (cytoplasmic) ring located beneath the basal body and composed of the three switch proteins. The C ring is filled with FlgE subunits, acting in the capacity of a measuring cup and then exported and assembled. Therefore, along the model, the length of the hook is determined by the capacity of the C ring. The two hook-associated proteins HAP1 and HAP3 are located at the hook–filament junction. The basal body consists of a rod, several rings, the Mot proteins, the switch complex and the flagellum-specific export apparatus. The rings anchor the flagellum to the cytoplasmic membrane (MS ring), the peptidoglycan (P ring) and the outer membrane (L ring) in Gram-negative bacteria. The flagella of Gram-positive bacteria are anchored only through the MS ring in the cytoplasmic membrane. The three switch proteins FliG, FliM and FliN allow the flagellum to switch rotation from clockwise to counterclockwise. This allows *E. coli* cells to chance direction of swimming in response to attractants or repellents sensed

by a complex chemotaxis system where the phosphorylated CheY protein contacts FliM (see Section 6.2.9). Assembly of the whole flagellum occurs through the type III secretion pathway (see Section 8.6.3).

Salmonella cells are approximately 0.73 μm wide and 1.4 μm long. Each cell possesses about five flagella and these flagella are arranged peritrichously (lateral) around the cell. Each cell rotates its flagella at ~200 revolutions s^{-1} and swims at speeds of ~55 μm s^{-1}. This is a propulsion rate of many cell body lengths per second. *E. coli* cells are similar: ~270 revolutions s^{-1} and 36 μm s^{-1}. In comparison, the marine bacterium *V. alginolyticus*, whose cell size is similar to the cell size of *E. coli* and *Salmonella*, swims at speeds as fast as 116 μm s^{-1} and its flagellar rotation rate is ~1000 revolutions s^{-1}.

High viscosity generally impedes flagellar rotations. As a consequence, bacteria implement additional strategies to maximize movement in viscous conditions and on surfaces, including the number of flagella and their arrangement. Many peritrichous bacteria upregulate the number of flagella and alter extracellular components such as polysaccharide and surfactant production to enable movement on surfaces, called swarming. Other bacteria induce completely new, alternative flagellar systems in response to growth in viscous environments and on surfaces. They are polarly flagellated when they grow in liquid medium and have mixed (polar and peritrichous) flagella when they are grown on surfaces. One example is the above mentioned *V. alginolyticus*, where the peritrichous flagella enable effective motility in highly viscous environments (20 μm s^{-1} in a ~200-cp environment). *B. subtilis* possesses two types of flagellar motors driven by different energy sources (the sodium and proton motive forces) but only one set of flagellar genes.

Pili (Fimbriae)

Pili (singular: pilus), also called fimbriae, are hair-like appendages built by protein subunits called pilin or fimbrin and usually extend 1–2 μm from the surface of Gram-negative bacteria, with a diameter ranging from 2 nm to 8 nm. They function in bacterial cell-to-cell interactions, adhesion to specific receptors of host cells, either uptake or transfer of genetic material and twitching motility, a form of locomotion that is powered by extension and retraction of the pilus filament, and they provide receptors for bacteriophages. Pili can be classified on the basis of their physical properties, antigenic determinants or adhesion characteristics. Class I and IV pili represent the two best characterized model systems and are ubiquitously present among Gram-negative pathogens. Pili are assembled by several distinct pathways and we will explain the biogenesis of F-pili, P-pili, type 4 pili, curli and T-pili.

F-pilus Biogenesis

The F-pilus with a length of about 1 μm is encoded by the *tra* genes located on the 100-kb F factor. It has a specialized tip structure, the nature of which remains unknown, and a helical array of subunits of 8 nm diameter with a 2-nm lumen. The

tip is involved in the recognition of a suitable recipient cell as well as F-specific filamentous phages such as M13, f1 and fd. RNA phages including R17 and Qβ attach to the sides of the pilus. The F-pilus is composed of F-pilin encoded by *traA* and translated as a 121-amino-acid propilin polypeptide. Correct insertion of the propilin into the membrane requires TraQ, an inner membrane protein. Propilin is processed by the signal peptidase (*lepB*) to a 77-amino-acid intermediate that is acetylated at its N-terminus by TraX. The mature F-pilin accumulates in the inner membrane as a pool of ~100 000 subunits. The F-pilus is assembled from this pool by eleven *tra* gene products (TraL, -E, -K, -B, -V, -C, -W, -F, -H, -G, TrbC). While TraA, -L, -E, -K, -D and -G are involved in assembling a structure at the cell surface, thought to be the pilus tip, TraB, -V, -W, -F and -H are involved in pilus extension, which is stabilized by TraP. TraU appears to be involved in DNA transfer and TrbI in pilus retraction after contact with the recipient cell.

P-pilus Biogenesis

Biogenesis of this superfamily of more than 30 pili requires outer membrane proteins known as *ushers* working together with periplasmic chaperones. P-pili represent the prototype organelle assembled by the chaperone–usher pathway, are expressed by many strains of uropathogenic *E. coli*, have been shown to be required for the establishment of pyelonephritis and are encoded by the *pap* (pyelonephritis associated pili) operon. Each pilus consists of a thick, rigid rod with a thinner, more flexible tip fibrillum at its distal end. The rod consists of PapA subunits arranged to form a hollow, right-handed helical structure 6.8 nm in diameter, while the tip fibrillum is composed mainly of repeating subunits of the PapE protein arranged in an open-helical fiber 2 nm in diameter. The PapG adhesin is situated at the distal end of the tip fibrillum and binds to Galα(1–4)Gal epitopes present in the globo series of glycolipids found in the human kidney. The PapF and PapK subunits are present in very low copy number in the pilus and connect the tip fibrillum to PapG. Assembly of the P-pilus involves two dedicated proteins, the *periplasmic chaperone* PapD and the outer membrane usher protein PapC and is described under Section 7.2.2.

Type 4 Pili Biogenesis

This pathway for assembling adhesive pili is exemplified by type 4 or bundle-forming pili expressed by a number of Gram-negative bacteria. These long, polarly localized pili are responsible for twitching motility and bacterial aggregation. Assembly requires 14 or more components, which are thought to organize into a supramolecular structure involved in secretion. Translocation of type 4 pili to the cell surface requires an outer membrane protein termed *secretin* forming large oligomeric rings with apparent central pores. The secretin plays a role analogous to that of the usher protein.

Curli Biogenesis

The third pathway assembles fibers known as *curli* or thin aggregative pili. Curli are thin filaments expressed by *E. coli* and *Salmonella* that can assemble extracellular amyloid fibers 4–6 nm wide. These fibers form a tangled extracellular matrix that connects several neighboring bacterial cells into small groups (Fig. 1.6).

Curli fibers are thin aggregation surface fimbriae, which are involved in cell–cell attachment and adhesion to extracellular matrices. They are also able to bind host proteins and might influence the host immune responses. Curli fibers resist protease digestion and remain insoluble when boiled in 1% sodium dodecyl sulfate. At least five proteins in *E. coli* are involved in assembling curli on the cell surface (see Section 8.6.2). Regulation of transcription of both operons is complex and is subject to different environmental cues, including osmolarity, temperature, growth phase and protein aggregation in the periplasm. The *csgBA* operon codes for the two homologous proteins, CsgA and CsgB, that are secreted into the extracellular environment. While the 13-kDa CsgA protein is the major component of the curli fiber, CsgB, the minor curli subunit, is required for CsgA polymerization at the cell surface. It has been suggested that CsgB induces a conformational change in CsgA that nucleates its assembly into fibers. CsgA can be purified in a soluble, unassembled state, which, after prolonged incubation, spontaneously as-

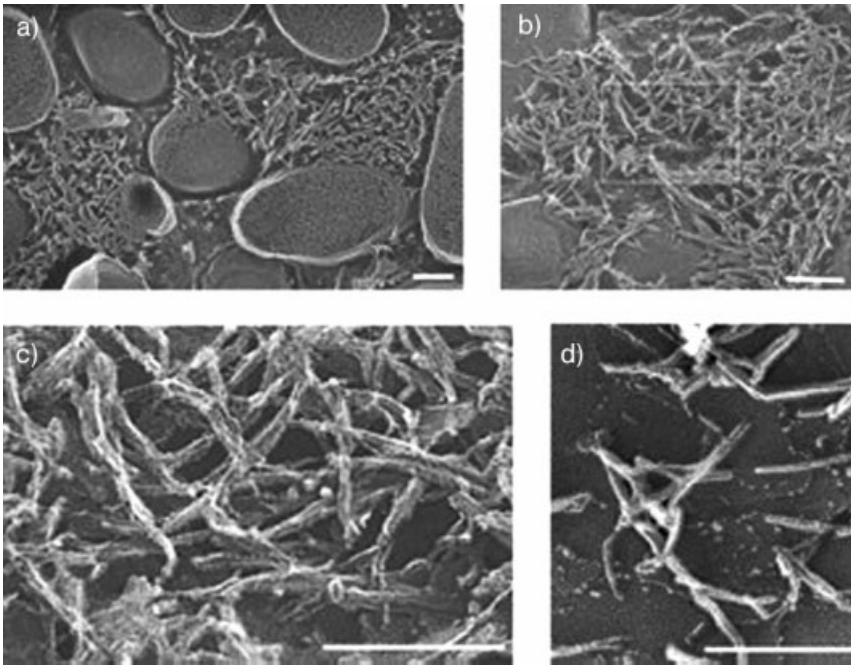


Fig. 1.6 High resolution deep-etch EM micrograph of curliated *E. coli cells*. (A–C) The curli fiber meshwork surrounding *E. coli cells*. (D) Purified CsgA major curli can form amyloid fibers by *in vitro* polymerization. M.R. Chapman, et al. **2003**, *ASM News* 69, 121–126; Fig. 1.

sembles into amyloid fibers. Many amyloid proteins assemble by nucleation-dependent polymerization, where the rate-limiting step is nucleus formation.

Why do bacteria produce amyloid fibers? Curli proteins are important mediators of biofilm formation and are able to bind to eukaryotic extracellular matrix proteins, such as fibronection and laminin. In addition, human macrophages recognize and respond better to curliated bacteria than to those not producing this extracellular matrix. Bacterial amyloids could also play a direct role in certain human neurodegenerative diseases by inducing host proteins to form pathogenic amyloid fibers. Such a cross-species nucleation has been demonstrated among prions from different genera of yeast.

T-Pilus Biogenesis

A. tumefaciens is able to induce tumors in plants (see Section 10.2.1), and pili (the T-pili) play a key role in virulence. One T-pilus is formed at one end of the cell when *A. tumefaciens* cells perceive phenolic compounds that signal the expression of virulence genes (see Section 6.2.3) which are located on a plasmid termed the Ti-plasmid (tumor-inducing). The products of the *virB* operon, comprising a total of 11 genes, are essential for the biogenesis of the T-pilus, which is mainly generated at one end of the bacterial cell. T-pili do not retract but wind into compact coils which presumably brings the bacterium and host cell into close proximity. In addition to the T-pilus, a number of flagella and common pili of 3–4 nm diameter are produced, arranged circumthecally and involved in motility. The common pili are encoded by the chromosome and their function remains elusive. The T-pilus serves as a conduit with its 2-nm width lumen for folded pilus subunits, the proteins VirB5 and VirE2 and the single-stranded T-DNA-VirD2 complex.

The T-pilin, 74 amino acids in length and product of the *virB2* gene, is ligated between its N- and C-termini by a peptide bond, forming a cyclic peptide (see Section 6.5.5) which is highly resistant to various chemical treatments including glycerol; and sodium dodecyl sulfate fails to dissociate T-pili. The remaining ten genes of the *virB* operon are thought to make up the T-DNA transmembrane transport apparatus, as shown in Fig. 1.7. This transport apparatus appears to also be used to deliver T-pilin subunits across the membranes to the outside of the cell where they assembly into the pilus. While VirB6, VirB7, VirB8, VirB9 and VirB10 are components of the transporter, VirB4 and VirB11 possess ATPase activity, which could be used to promote the transport of the T-DNA complex and the T-pilin subunits. VirB1*, the processed form of VirB1, is present in the culture supernatant and is thought to aid transiently in the translocation of T-pilin subunits. It could be released when T-pilin subunits reach the outer bacterial cell surface. VirB5 cofractionates as a minor component with T-pili preparations and could stabilize T-pilin multimerization and assembly into the T-pilus filament. A potential prerequisite for T-pilus formation is the hydrolysis of peptidoglycan. The VirB1 protein, a transglycosidase, might be involved in peptidoglycan hydrolysis, allowing the T-pilus to assemble and penetrate through the peptidoglycan layer.

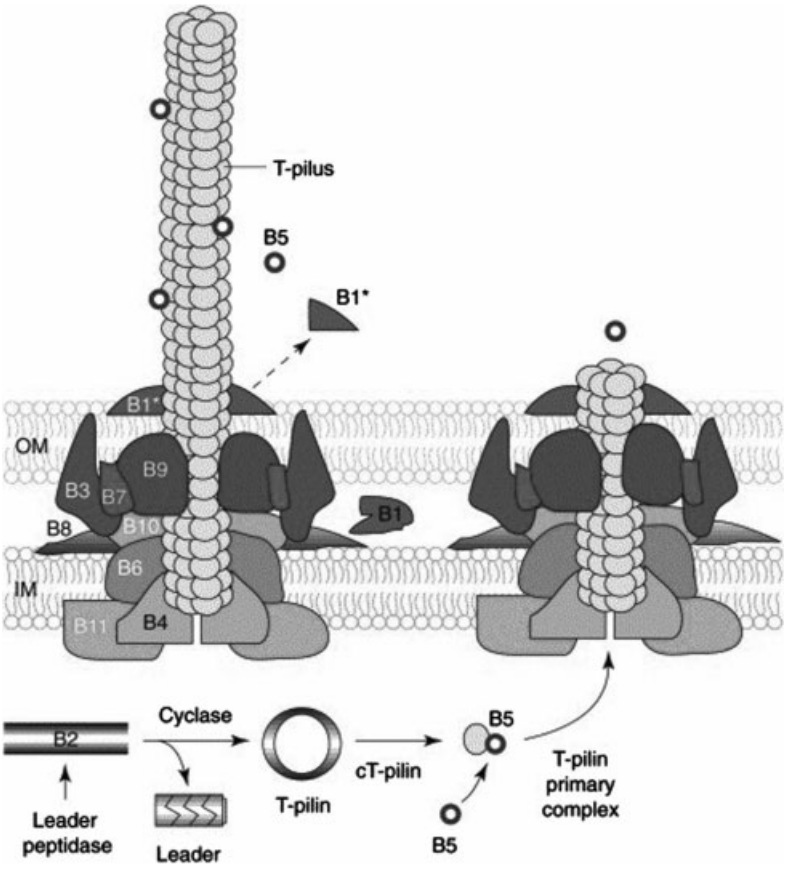


Fig. 1.7 T-pilus biogenesis. A transmembrane transport apparatus is composed of VirB proteins that provide a scaffold for the assembly of T-pilin subunits. The pilus itself consists of circular T-pilin subunits with some attached VirB5 subunits. E. Lai, C.I. Kado **2000**, *Trends Microbiol.* 8, 361–369; Fig. 5.

2
Organization of the Bacterial Chromosome

In most cases, eubacteria contain a single *circular* chromosome between 500 and 10000 kb in size. But there are bacterial species with *linear* chromosomes (*Borrelia*, *Streptomyces*) and in rare cases bacteria harbor *one circular* and *one linear* chromosome (*Agrobacterium*). In addition, many bacterial cells contain one or more *plasmids* and the chromosomes normally carry *prophages*. An extreme example is represented by the food-borne pathogen *E. coli* O157:H7 Sakai, which contains 18 prophages that account for 16% of its genome. *Insertion sequences* and *transposons* are present in virtually all prokaryotes. All the genetic information present within a single cell is termed the *genome*. The genomes of bacteria are remarkably fluid, since a substantial portion of the genes has not been inherited from parental cells. Rather, it has been acquired horizontally by a process called *h*orizontal *g*ene *t*ransfer (HGT), mediated by the genetic recombination mechanisms transduction, transformation and conjugation. The acquisition of new genes by HGT allows a sudden adaptation to new environments (see Section 10.5). The length of the bacterial chromosome can exceed that of the cell by a factor of 100 and more, and to fit into the cell, the chromosomes have to become compacted by the introduction of *superhelical coils* and by the binding of *histone-like proteins*. About 90% of a typical chromosome codes for proteins and structurals RNAs, while the remaining 10% carry *repetitive sequences* and DNA sequences involved in gene regulation.

2.1
Structure of the Chromosomes

While most bacterial chromosomes are circular, some species contain linear chromosomes, such as the Actinomycetes and the spirochete *B. burgdorferi*, in which the first linear bacterial chromosome was detected. Also, a few species such as *A. tumefaciens* even harbor one circular and one linear chromosome. Our current knowledge on the distribution of circular and linear chromosomes in bacteria is presented in Fig. 2.1.

This figure clearly demonstrates that circular chromosomes are preferred; and linear chromosomes may have been arisen by the linearization of circular chromosomes independently in several bacterial lineages. Chromosome linearization

Dynamics of the Bacterial Chromosome. Structure and Function. W. Schumann
Copyright © 2006 Wiley-VCH Verlag GmbH & Co. KGaA
ISBN 3-527-30496-7

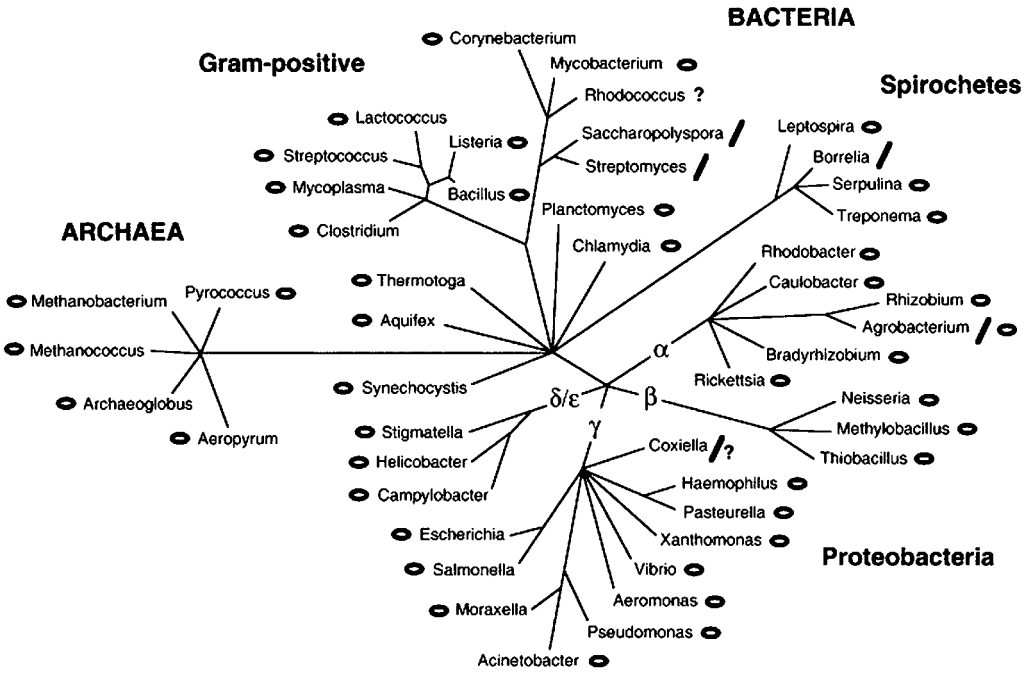


Fig. 2.1 Distribution of circular and linear chromosomes among major procaryotic groups. J.-N. Volff **2000**, *FEMS Microbiol. Lett.* 186, 143; Fig. 1.

produces interesting consequences for DNA replication termination (see Section 3.1.7), replicative transposition and conjugation.

Streptomyces bacteria live in the soil and must adapt to a wide range of terrestial environments. So far, the genome of two different species have been sequenced, namely *S. coelicor* and *S. avermitlis*. The central portion of the linear 8.7-Mb chromosome of *S. coelicor* M145 encodes many genes likely to be essential for growth, including those involved in cell division and DNA replication. In contrast, the chromosome arms contain many genes with nonessential functions, such as those of hydrolytic and secondary metabolism enzymes, which have auxiliary function under specialized conditions. The internal portions of the *S. avermitlis* and *S. coelicor* M145 genomes are highly conserved. This observation may suggest that the 'core' versus 'auxiliary' organization is a general feature of *Streptomycetes* genomes. The ends of *Streptomycetes* chromosomes are highly dynamic, exhibiting a great variation in length. The inverted repeat sequences identified at the chromosome ends (termed *t*erminal *i*nverted *r*epeats; TIRs) range from 175 bp for *S. avermitlis* to 850 kb for *S. ambofaciens*. Recently, several used laboratory strains of *S. coelicor* A3(2) were found to have 1.06 Mb TIRs and are the longest reported thus far for an actinomycete chromosome. The ends of *Streptomycetes* chromosomes also undergo large deletions at frequencies between 0.1% and 1.0%. Deletions are often accompanied by circularization of the chromosome and amplification of subtelomeric DNA.

Chromosomes are dynamic in the sense that DNA from outside sources can be introduced in several ways. Approximately half of the sequenced bacterial genomes have readily identifiable prophage sequences. Furthermore, incoming plasmids can also integrate into bacterial genomes, and any mobile element they carry can accumulate over times, sometimes reaching very large numbers (some strains of *Shigella* contain hundreds of copies of IS*1*). Surveys of microbial genomes reveal large amounts of laterally transferred DNA in the majority of bacteria, constituting up to 20% of their genomes. Why is the chromosome size not ever-increasing? While the genomes of characterized bacteria vary between 500 kb and 10000 kb in size, the genome sizes of natural variants of *E. coli* vary far less, measuring 4950±250 kb. There is even less variability in the chromosomes of the intracellular parasite *Buchnera* (630–643 kb) and the obligate parasite *Borrelia* (903–922 kb). Thus, the constant influx of DNA by lateral transfer appears to be compensated by persistent gene loss, resulting in an approximately steady-state genome size.

2.2 Principles to Compact the Bacterial Chromosome

The contour length of the chromosomal DNA is approximately 1000 times the length of the chromosomal DNA that it occupies. Therefore, the chromosome has to be organized into a compact structure, the *nucleoid*. The nucleoid consists of multiple topologically independent, negatively supercoiled loop domains (average size of just 10 kb in *E. coli* that are randomly positioned around the chromosome). For the *E. coli* nucleoid, a volume of ~0.2 μm^3 has been calculated (20% of the total volume). A search for *domainins*, proteins involved in maintaining domains, has been initiated; and two candidate proteins, H-NS and DksA (a suppressor of *dnaK* and *mukB* mutations) have been identified so far. The level of supercoiling in bacteria is mostly maintained by the coordinate action of two activities: DNA gyrase, which introduces negative supercoils, and DNA topoisomerase I, which removes negative supercoils. Besides constrained supercoiling bacterial chromosomal DNA condensation involves binding of several proteins dominated by so-called *histone-like proteins*. Histone-like proteins structure the bacterial nucleoid and also exert regulatory functions on transcription, recombination and DNA replication. Furthermore, they are involved in modifications of the DNA geometry since they recognize, create or even exacerbate bends in DNA and influence the level of DNA supercoiling.

2.2.1 Superhelicity

In bacteria, all circular DNA molecules are negatively supercoiled. This means that the DNA is twisted in the opposite direction to the WatsonCrick helix, thereby creating underwinds. This twisting produces loops in the chromosome, causing it

be condensed into a smaller space. Negative superhelicity of DNA is necessary for both DNA replication and transcription and is essential for cell viability. Therefore, negative superhelicity is tightly maintained.

In *E. coli*, superhelicity is maintained through the balanced actions of topoisomerase I (topo I), DNA gyrase and topo IV. Whereas DNA gyrase introduces negative supercoils, topo I relaxes negative supercoils and topo IV relaxes both positive and negative supercoils. DNA gyrase and topo IV are type-2 topoisomerases that change the linking number in steps of two and require ATP binding and hydrolysis for their functions. Type-2 topoisomerases introduce a transient double-strand break and pass a second duplex segment through the break, followed by ligation of the break. In contrast, topo I belongs to type-1 isomerases which change the linking number in steps of one and do not require ATP. They cut only one strand and pass the other strand through the break before resealing the cut.

Structurally, gyrase acts as an $\alpha_2\beta_2$ tetramer, where the monomers are encoded by the *gyrA* and *gyrB* genes. Approximately 130 bp of DNA are wrapped around the enyzme and gyrase acts by catalyzing the formation of a 4-bp staggered double-strand break in a central region of the wrapped DNA (the 'gate' or G segment). It then passes a separate region of the wrapped DNA (the 'transporter' or T segment) through the G segment break. The introduction of supercoils occurs at the expense of ATP hydrolysis. In the absence of a high-energy cofactor, DNA gyrase is capable of relaxing DNA. While the GyrB protein forms the site of ATP binding and hydrolysis, the GyrA subunit is involved in catalyzing the breakage and reunion of the G segment. Each Tyr122 residue of the *E. coli* GyrA forms a covalent link to the 5′ ends of the double-strand break during the strand passage event. The level of DNA supercoiling is regulated by a complex homeostatic control. Transcription of *topA*, which encodes Topo I, increases when the level is high, whereas transcription of the *gyr* genes increases when the level is low. It is well established that changes in the level of DNA supercoiling influence the activity of many promoters and that environmental variations alter this level.

2.2.2 Histone-like Proteins in *E. coli*

E. coli contains at least 240 different molecular species of DNA-binding proteins. A group of several abundant proteins constitutes the compact nucleoid architecture by condensing the genome DNA, which is about 1000-fold longer than the cell size. These are designated as histone-like or nucleoid-associated proteins and are thought to be involved in DNA compaction. The proposed function of these proteins has been derived from copurification with isolated nucleoids. In addition, their role in DNA compaction is based on superficial similarities with eukaryotic histones, such as small size, DNA-binding capacity, abundance and basicity.

In *E. coli*, the major histone-like proteins HU (*h*eat-*u*nstable nucleoid protein), IHF (*i*ntegration *h*ost *f*actor), H-NS (*h*istone-like *n*ucleoid *s*tructuring protein), Fis (*f*actor for *i*nversion *s*timulation), StpA (*s*uppressor of *td* mutant *p*henotype *A*) and one stationary-phase-specific DNA-binding protein Dps (*D*NA binding *p*rotein

from starved cells) act directly or indirectly, and often in combination, to compact the chromosome and to control a wide variety of genes. These genes are integral for cell viability, such as those involved in protein synthesis and carbon metabolism. In addition, they regulate genes whose expression varies in response to environmental stimuli, like temperature and osmolarity.

Heat-unstable Nucleoid Protein

HU is the major component of the bacterial nucleoid. It affects the overall nucleotide structure and topology but also participates in specific gene regulation, DNA recombination and DNA repair. Additionally, HU is required for optimal survival during prolonged starvation and has been identified in all eubacteria analyzed so far. In most bacteria, it exists as a 18-kDa homodimer, but in *E. coli* (and other Enterobacteriaceae) HU is predominantly found as a heterodimer composed of the two homologous subunits, HUα and HUβ, encoded by the *hupA* and *hupB* genes. During growth, the HUα2 homodimer is abundant, whereas during the late exponential growth phase, *hupB* is induced and HUαβ heterodimers are formed in *E. coli*. HU interacts with DNA in a sequence-independent manner and preferentially recognizes bent or kinked DNA or DNA in cruciform structures. Binding to DNA probably involves its two arms being inserted into the minor groove. Upon binding, HU bends DNA and wraps it into nucleosome-like structures. Besides affecting the overall nucleoid structure and topology, HU-induced DNA distortions influence a variety of activities, either by stimulating the binding of regulatory proteins to their target (exemplified by the *lac* operon) or by inducing the formation of a loop which stabilizes the interactions between regulatory proteins (for instance, the *gal* operon). HU is even able to strongly bind to the mRNA of the *rpoS* gene, stimulating its translation.

Integration Host Factor

IHF is a basic heterodimer composed by two distinct subunits coded for by the unlinked *himA* and *himD* genes. Unlike HU, IHF binding to DNA exhibits sequence specificity. IHF sites are approximately 30 bp long and contain a 13-bp consensus sequence (WATCAANNNNTTR). Upon binding, IHF strongly bends DNA by angles up to 160°. This activity is crucial to its biological role that facilitates the construction of intricate DNA–protein complexes. IHF directly or indirectly affects the transcription of over 100 *E. coli* genes coding for disparate functions.

Factor for Inversion Stimulation

Fis is another basic protein encoded by the *fis* gene and conserved in most enteric bacteria. Fis is active as homodimer that binds to sites that display a highly degenerate consensus sequence (GNYAWWWTRNC) and induce DNA bending up to 90°. Fis, originally described for its crucial role in site-specific inversion of the G

segment of bacteriophage Mu DNA (see Section 4.2.3), participates in other fundamental cellular processes, such as chromosomal replication, DNA transposition and the stimulation of stable RNA synthesis. Fis is produced in high amounts during early exponential phases of cell growth and the dramatic variation of its intracellular concentration is consistent with its role as transcriptional modulator. Currently, Fis is recognized as a global modulator of metabolism, coupling cellular physiology to chromosome topology. Fis represses transcription of both DNA gyrase subunits, counteracting in this way the generation of excessive superhelical tension during an intense transcriptional phase. Fis is assumed to interact with the α-CTD of the α subunit of the RNA polymerase (see Section 6.2.1), though its action may also depend on the ability to bend DNA. Fis is involved in the regulation of a subset of genes by direct control of transcription initiation as well as by possible indirect effects on some RpoS-regulated genes.

Histone-like Nucleoid Structuring Protein

The highly abundant 15.6-kDa H-NS (>20000 copies per cell) exhibits a global neutral pI but harbors many patches of charged amino acids. It is encoded by the *hns* gene which binds to DNA with no obvious sequence specificity, preferentially recognizing intrinsically curved DNA and AT-rich sequences, and is able to induce bending of noncurved DNA and negative supercoiling. Binding of H-NS generally induces strong condensation of DNA, leading to the suggestion that H-NS is involved in the organization and compaction of the bacterial nucleoid (Fig. 2.2). H-NS is able to form homodimers in solution. Extending from the initial binding site, H-NS forms oligomeric structures on the DNA, which prevent binding of RNA polymerase or which trap RNA polymerase at the promoter. Among the *E. coli* histone-like proteins, H-NS is by far the most important global modulator. H-NS negatively controls its own promoter, antagonizing Fis-mediated activation. Besides regulating genes coding for housekeeping functions, H-NS

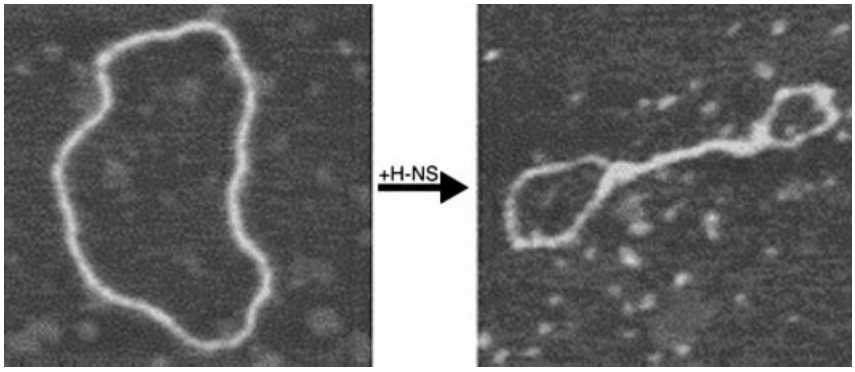


Fig. 2.2 H-NS compacts DNA molecules. Relaxed pUC19 DNA (left panel) is incubated with H-NS leading to compaction (right panel). R.T. Dame, N. Goosen **2002**, *FEBS Lett.* 529, 151–156; Fig. 1.

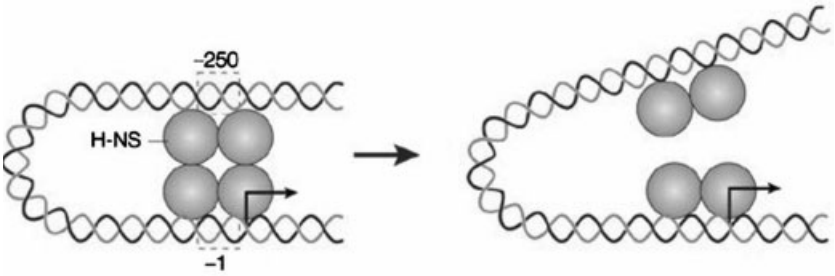


Fig. 2.3 H-NS-mediated repression of the *virF* promoter of *Shigella flexneri*. At low temperature, H-NS binds to two different sites located at –250 and –1 relative to the transcriptional start site of the *virF* promoter and both complexes interact with each other to repress transcription initiation (left panel). When the temperature increases (after infection of a mammalian host), both complexes dissociate allowing transcription initiation. C.J. Dorman **2004**, *Nat. Rev. Microbiol.* 2, 391–400; Fig. 2a.

affects gene networks involved in bacterial response to multiple environmental conditions. It directly or indirectly affects expression of approximately 5% of the *E. coli* gene, many of which play a role in response to environmental stimuli.

Dimerization of the 136-amino-acid H-NS protein requires the N-terminal 45 amino acids, whereas the C-terminus encompasses the DNA-binding domain. The central region of H-NS, which is unstructured in solution, is required for oligomerization on the DNA. H-NS effects transcription mainly by direct binding as a repressor, but also acting as a silencer of extended chromosomal regions. How the activity of H-NS is modulated and how the high specificity of repression by H-NS is achieved are largely open questions. The H-NS mediated repression of several genes is relieved by antagonistic positive transcription factors. Examples are the *hns* gene itself, which is activated by the DNA-bending protein Fis, and the virulence gene *virB* by the transcriptional activator VirF. *In vitro* experiments show the formation of higher-order structures of H-NS on the DNA is reduced at elevated temperatures (37 °C vs 25 °C) and by increased salt concentrations (Fig. 2.3). These physical properties of H-NS may explain the induction of several H-NS repressed virulence genes at 37 °C and the *proU* operon of *E. coli* at high osmolarity.

Suppressor of *td* Mutant Phenotype A

StpA is a basic protein sharing very high amino acid sequence homology with H-NS. In both proteins, a flexible linker region connects the C-terminal (DNA binding) to the N-terminal domain (oligomerization). Originally identified for its ability to stimulate splicing of the T4 group I *td* intron, StpA can replace H-NS in the control of some genes and is able to form active dimers with H-NS. StpA acts mainly as an RNA chaperone (see Section 6.3.6), but is also able to bind curved DNA and to constrain supercoils.

DNA Binding Protein From Starved Cells

Dps is among the most abundant proteins in stationary-phase *E. coli* cells. As cells undergo the transition from exponential-phase growth to the stationary phase, Dps is induced in a σ^S-dependent manner, reaching up to 200000 molecules per cell. The *dps* gene is also regulated by OxyR and σ^{70} in response to oxidative stress during the exponential phase. The protein has nonspecific DNA-binding activities, affects the expression of at least three dozen genes, and plays a role in various stress responses. Dps protects DNA effectively, both *in vitro* and *in vivo*, against oxidative cleavage, nucleases, UV light and thermal shock. The protein's active form is a hollow, spherical dodecamer with an outer diameter of 90 Å and interior core diameter of 45 Å. During the stationary phase, Dps binds to chromosomal DNA, forming a highly ordered and stable nucleoprotein complex called a biocrystal. This biocrystalline form is distinct from the exponential-phase configuration of the nucleoid. Biocrystal formation contributes to the ability of Dps to affect gene expression and protects chromosomal DNA. However, the actual mode of the Dps-DNA interaction has not been fully elucidated. To explain Dps-DNA interaction, two models have been proposed. One model assumes that, in the absence of canonical DNA-binding motifs, interaction is mediated through bridges formed by Mg^{2+} in a particular range of cation concentrations. The second model takes into consideration a contribution of the disordered, lysine-containing N-terminus (three lysine residues, at positions 5, 8, 10). In the hexagonal Dps crystal lattice, three adjacent dodecamers define holes that are lined by the lysine-rich N-termini. In the Dps-DNA complex, DNA is assumed to be threaded through such holes and to interact with the lysine residues. The second model is in accordance with the observation that members of the Dps family that do not possess a positively charged N-terminus do not bind DNA, based on agarose gel mobility assays.

2.3 Organization of the Bacterial Chromosome into Genes and Repetitive Sequences

2.3.1 Genes

Bacterial genes are short, averaging only ~1000 nucleotides in each of the sequenced genomes. Many annotated genes are no more than putative *o*pen *r*eading *f*rames (ORFs), as recognized by in-frame start and stop codons and preceded by a ShineDalgarno sequence. Among the shortest ORFs within the *E. coli* genome are well recognized leader peptides of only 1530 amino acids; and several ribosomal proteins are less than 100 amino acids in length. Genes are organized into transcriptional units, and in most cases, transcriptional units (operons) contain more than one gene. Transcriptional units are separated by intergenic regions which are relatively short in prokaryotic genomes compared to those in eukaryotic genomes. There are three types of transcriptional units with respect to the direction

of transcription: (1) unidirectional, (2) convergent and (3) divergent. Spacers between unidirectional genes represent a mixture of inter- and intra-operonic spacers and may include both a terminator for the upstream gene and a promoter and additional signals for the downstream gene(s). Spacers between convergent genes contain exclusively terminators, and spacers between divergent genes have only promoters and other upstream transcriptional signals.

Bacterial genomes may also contain a variable number of *pseudogenes*. Pseudogenes are ORFs encoding dysfunctional proteins with high homology to known protein-coding genes. Pseudogenes can occupy 27% of the genome of the obligate parasitic bacterium *M. leprae*. A systematic investigation of the whole genome sequences of *E. coli* strains K-12 and O157, which diverged approximately 4.5 million years ago, identified 95 and 101 pseudogene candidates, respectively. The assigned three-dimensional structures suggest that most of the encoded proteins are unable to fold properly and thus are dysfunctional.

2.3.2 Repetitive Sequences

DNA repeats are defined as sequences sharing extensive similarity with other sequences of the same genome. It is usually assumed that they arise by successive duplications. Once repeats are created, they can be targeted by the recombination apparatus and be subject to deletion or inversion depending on their relative orientations. Usually, repeats in bacteria are divided into two subclasses: low complexity repeats without coding capacity and longer repeats with coding capacity. The first category is constituted of small oligonucleotides (examples are the chi sequence of *E. coli* with a length of 8 bp and the 10-bp uptake sequence of *N. gonorrhoeae*) up to the 137-bp YPAL2 element (see below). Repetitive elements with coding capacity include the *rrn* operons and IS element, to mention two of them.

2.3.2.1 Repetitive Sequences Without Coding Capacity

Repetitive elements without coding capacities can be classified according to their size and their number in the genome. The shortest are the chi sequences of *E. coli* and other bacterial species which play an important role during homologous recombination (see Section 4.1.1.1). The *E. coli* K-12 chromosome contains a total of 16384 copies of chi. Other short repetitive elements are the so-called uptake sequences present in *H. influenzae* (AAGTGCGGTCA or its inverse complement TGACCGCACTT) and *N. gonorrhoeae* (GCCGTCTGAA), which play a role during transformation (see Section 10.1). The *V. cholerae* repeats (VCR) are 121-bp to 126-bp sequences of imperfect dyad symmetry, occur in more than 150 copies and act as recombination sites of the super-integron (see Section 4.2.5).

In addition, there are bacterial palindromic repeats present in enterobacteria, less than 200 bp long, extragenic, transcribed and with the ability to form stem-loop structures. Their level of repetition is between ten and 500 copies. These repeats can be classified into simple and composite repeats. First, we will describe

four simple (ERIC, RSA, YPAL1, YPAL2) and then four composite repeats (29-bp repeat, BoxC, BOCE, BIME).

ERIC and RSA The ERIC (*e*nterobacterial *r*epetitive *i*ntergenic *c*onsensus) and RSA elements are 127-bp and 151-bp long, respectively. While six copies of RSA and 21 copies of ERIC are found in the *E. coli* K-12 chromosome, they are found in a wide array of enterobacteria with different levels of repetition. In some species, ERIC and RSA elements are interrupted by a conserved palindromic intervening sequence of 62 bp or 68 bp, able to form a stable stem-loop structure. The function of both elements are unknown; and it has been speculated that the secondary structures on RNA molecules interfere with translation or to protect mRNA from exoribonuclease digestion. Furthermore, these elements could serve as interaction sites for proteins.

YPAL1 and YPAL2 These two palindromic elements have been detected in the *Y. pestis* genome where they occur in 65 and 15 copies, respectively. YPAL1 is 167 bp long, while YPAL2 has a length of 137 bp.

The 29-bp Repeat The 29-bp repeats were first identified next to the *iap* gene (coding for an alkaline phosphatase) of *E. coli*. Fourteen repeats are separated by 32-bp or 33-bp spacers. Two additional groups of the 29-bp repeat are present about 24 kb downstream of the *iap* gene, where the first group contains two repeats and the second group has seven, each time associated with 32-bp spacers. The 29-bp repeat is also present on the chromosome on *S. typhimurium* and *S. typhi*, but not on that of *K. pneumoniae*. The function of this repeat remains elusive.

BoxC and BOCE BoxC sequences are 56-bp imperfect palindromes with a pyrimidine-rich 5′ end (called the tail) and a purine-rich 3′ end (head). The *E. coli* K-12 chromosome contains a total of 22 extragenic regions with either one or two BoxC elements, totalling 32 BoxCs. When two BoxCs are present in the same region, they are either in an inverted or tail-to-tail orientation. When two BoxC elements are found in the tail-to-tail orientation, an 11-bp conserved sequence motif called 'D' is found between the two elements, and a 25-bp accessory motif called 'E' can be found flanking the head of one element. This composite element has been called BOCE (for *B*oxC *c*omposite *e*lement). *E. coli* K-12 contains a total of 13 BOCE elements. Similar elements have been found in *S. sonnei*, *K. pneumoniae* and *E. cloacae*. The functional significance of BoxC and BOCE elements is not known. The dyad symmetry and their alternate orientation in BOCE suggest stable secondary structures to be formed which could stabilize mRNAs and/or be involved in the regulation of transcription. Alternatively or in addition, they could act as binding sites for proteins.

PU (REP) and BIME PU (*p*alindromic *u*nit) or REP (*r*epetitive *e*xtragenic *p*alindromic) elements are 40-bp imperfect palindromes of *E. coli*, *S. typhimurium* and *K. pneumoniae* chromosomes. On the *E. coli* K-12 chromosome, PU elements are

either found as single occurrences (42 copies) or associated with short conserved sequence motifs, to form a mosaic element termed BIME (*b*acterial *i*nterspersed *m*osaic *e*lement). The PU elements can be subdivided into two groups called Y and Z, according to the nucleotides at the seventh and 32nd positions of the original consensus sequence. PU elements can be combined with seven different motifs, where S and L are located between PU heads, l, s and r are located between PU tails and the motifs A and B can be found flanking the tail of one PU. Preferential association between PU elements and these motifs lead to the definition of two major BIME families, BIME-1 and BIME-2 (Fig. 2.4). Both families contain two copies of PU separating a central motif termed L and S, respectively, where the L sequence serves as a binding site for IHF. There are 61 BIME-1 and a total of 71 BIME-2 scattered on the *E. coli* K-12 chromosome. The BIMEs are not evenly distributed on the chromosome and their density is lower in the region of replication termination.

What is the function of the BIMEs? Some PU elements, termed PU*, serve as bidirectional terminators. When a BIME is located within an operon, the upstream gene often has a higher level of expression than the downstream gene(s). One prominent example is the *malEFG* operon which is transcribed as a 3.9-kb tricistronic transcript which is rapidly processed into a stable 1.3-kb *malE* product. The *malE* and *malFG* genes are separated by a BIME element which acts as a stabilizer of the mRNA by forming secondary structures, protecting the RNA against 3′→5′ exonuclease degradation. But BIMEs alone are not always able to increase mRNA stability when fused to other transcripts. Several proteins specifically interact with BIMEs. Pol I binds to PUs in a BIME-2 structure *in vitro* and DNA gyrase contacts PUs in BIME-2 *in vitro* and *in vivo*. As already mentioned, IHF interacts with the L motif in BIME-1.

S. Bachellier, et al. **1999**, Short palindromic repetitive DNA elements in enterobacteria: a survey, *Res. Microbiol.* 150, 627–639.

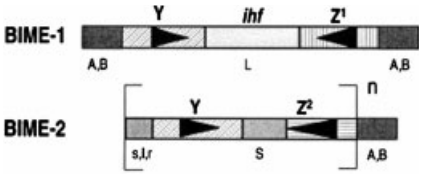


Fig. 2.4 Schematic representation of the composition of BIME-1 and BIME-2. BIME-1 is composed of two copies of PU (denoted as Y and Z1) separated by an L motif, a binding site for IHF, and flanked by either A or B motifs. BIME-2 also contains two PU elements (here denoted as Y and Z2) separating an S motif and either an l or s or r motif at the left end. This unit can be amplified and is flanked at the right end by either one A or one B sequence. O. Espéli, et al. **2001**, *J. Mol. Biol.* 314, 375.

2.3.2.2 Repetitive Sequences with Coding Capacity

***rrn* Loci** *E. coli* contains seven rRNA operons, each transcribed from tandem promoters, *rrn* P1 and *rrn* P2. Most rRNA transcription initiates from P1 promoters during rapid growth, while the P2 promoter accounts for most rRNA transcription at low growth rates and in prolonged stationary phase. Regulation of rRNA promoters results primarily from the action of the initiating NTP (iNTP) and the alarmone guanosine 5′-diphosphate 3'-diphosphate (ppGpp) that act on a kinetic intermediate, the open complex, whose lifetime is rate-limiting for rRNA transcription initiation. rRNA promoters are unusually sensitive to the concentrations of these two small molecules. An increase in the cellular concentration of NTP leads to a transient stabilization of the rRNA promoter open complex by the iNTP driving transcription initiation forward. ppGpp is synthesized primarily by the ribosome-associated RelA protein in response to uncharged tRNAs in the ribosome's A site. ppGpp binds to a site adjacent to, but not overlapping with, the active site of the RNA polymerase and decreases the lifetimes of all open complexes, inhibiting transcription from promoters in which this complex is intrinsically short-lived. In addition, the DksA protein is absolutely required for rRNA regulation. In the absence of this protein, rRNA promoters are unresponsive to changes in amino acid availability, growth rate, or growth phase. DksA binds to purified RNA polymerase, reduces open complex lifetime, inhibits rRNA promoter activity and amplifies the effects exerted by ppGpp and the iNTP on rRNA transcription.

***rhs* Loci** The Rhs (for *r*earrangement *h*ot *s*pot) elements are complex accessory DNA elements found in the genomes of many natural *E. coli* strains which share features and homologies. Independent *E. coli* strains may have multiple Rhs elements, but there are also strains known with none. Five Rhs elements are present in *E. coli* K-12 and make up 0.8% of its chromosome. There average length is about 8 kb and they are the largest repetitious sequences known in the K-12 genome. Rhs elements have assembled from several discrete components; and the prototype structure of an Rhs element is shown in Fig. 2.5.

The most prominent component of an Rhs element is the 3.7-kb, GC-rich Rhs core which provides homology for a frequent, unequal recombinational event (hence Rhs). The core contains a single ORF that extends the full length of the core and has the capacity to encode a 141-kDa protein. The ORF does not terminate at the end of the core, but extends into the adjacent AT-rich region for an additional 139–177 codons (called core-extension), depending on the element. Consequently, the elements are predicted to produce protein up to 160 kDa with highly conserved N-termini and variable C-termini separated by a nine-codon joint segment. The Rhs core protein contains a peptide motif (GxxxRYxYDxDxxGRL[I or T]) that is repeated 28 times. The predicted core protein sequence is similar to the sequence of a *B. subtilis* wall-associated protein, leading to the speculation that the core product is a surface protein with binding function. It may provide the cell with an advantage in a specific habitat such as a mammalian host. Some Rhs elements contain in addition the *vgr* (*v*aline-*g*lycine *r*epeats) gene characterized by a

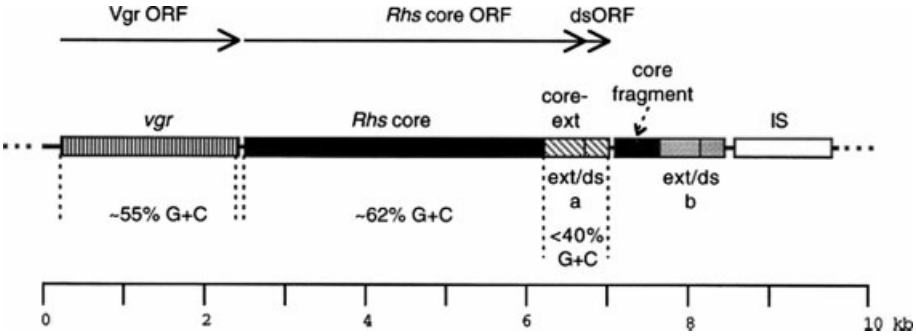


Fig. 2.5 Organization of a prototypical Rhs element. The most conserved region of the Rhs elements is the core sequence which is linked in some elements to the core-extension followed by one or more (mostly defective) IS elements. RhsE and RhsG contain in adddition the *vgr* gene. C.W. Hill, **1999**, *Res. Microbiol.* 150, 665–674; Fig. 4.

regularly repeated valine–glycine motif. The deduced hydrophilic protein may serve as ligand-binding protein. *vgr* homologs are also found in other Gram-negative bacteria, but they do not appear to be present in any Gram-positive organism yet sequenced. *P. aeruginosa* carries ten randomly distributed copies of *vgr* in its chromosome. The biological function of Rhs elements remains unknown. They are not essential for the cell, but they are widespread in natural *E. coli* populations, indicating that these elements are maintained by selective pressure.

IS Elements *I*nsertion (IS) elements are usually 700 bp to 2000 bp in length; and these genetically compact DNA sequences encode only functions involved in their translocation and transpose both within and between genomes. IS elements carry no selectable genes and have been identified only by their ability to generate mutants by insertional inactivation. To date, more than 600 such elements have been identified from 171 bacterial and archaeal species (www.is.biotoul.fr/is.html). IS elements are characterized by the following parameters:

- They contain short inverted repeats (IR) at their ends.
- They code for an enzyme called transposase catalyzing transposition.
- Few IS elements code for an additional function involved in regulation of transposition (e.g., IS*3*, see below).
- They cause duplication of target site sequences the length of which is specific for the element (214 bp).
- Most elements integrate into target sites without any specificity.
- IS elements are involved in DNA arrangements such as duplication, inversion and mobilization of neighboring genes (cotransposition).
- Upon integration into expressed genes, they cause an insertion mutation.
- Some elements can turn on gene expression by providing the 35 region recognized by the housekeeping promoter (e.g., IS*2*).
- IS elements can flank antibiotic resistance marker and thereby form *composite transposons*.

The IS*3* element consists of two consecutive and partially overlapping reading frames called *orfA* and *orfB*, preceded by a promoter. The reading frame of *orfB* is shifted –1 relative to the reading frame of *orfA*. Two different proteins can be synthesized, a smaller one encoded by *orfA* (regulates transcription of both *orf*s) and a frameshift product, the transposase. The ORFA protein (molecular mass about 11 kDa) contains a *h*elix-*t*urn-*h*elix (HTH) motif probably involved in sequence-specific binding to the terminal IRs and a C-terminal *l*eucine *z*ipper (LZ) motif involved in protein multimerization. The ORFB protein (35 kDa) contains the highly conserved amino acid triad DDE constituting part of the active site and is produced in very low amounts, mostly due to its unusual initiation codon AUU restricted in *E. coli* to the *infC* gene coding for the translation initiation factor IF3. ORFA acts as a transposition inhibitor and ORFB seems to enhance this inhibitory activity. The expression of *orfB* is translationally coupled to that of *orfA* by virtue of overlapping *orfA* termination and *orfB* initiation codons involving a –1 frameshift modulated by a potential downstream pseudoknot and a group of slippery codons (see Section 6.4.7.2). Frameshifting occurs with a frequency of about 15% and generates a protein that combines the HTH and the LZ motifs of ORFA and the DDE catalytic domain of ORFB, the active transposase.

J. Mahillon, et al. **1999**, IS elements as constituents of bacterial genomes, *Res. Microbiol.* 150, 675–587.

Bacterial Retrons A peculiar single-stranded satellite DNA was discovered in 1984, first in the Gram-negative soil bacterium *M. xanthus* and later in *S. aurantiaca*. Because it exists in a cell at a level of 500–700 copies per chromosome, it was designated as *m*ulticopy *s*ingle-stranded DNA (msDNA). The msDNA molecule of *M. xanthus* is composed of a single-stranded DNA of 162 bases with an attached RNA molecule of 77 nucleotides. The 5′ end of the DNA strand is linked to the 2′ OH group of a specific guanosine residue (referred to as the branching G residue) in the middle of the RNA strand, forming a 2′–5′ phosphodiester bond (Fig. 2.6A). Thus, msDNA is actually a complex of DNA, RNA and probably also protein.

The locus encoding the msDNA consists of three genes: *msr*, which encodes the RNA part of msDNA, *msd*, which encodes the DNA part, and the *ret* gene, for *re*verse *t*ranscriptase (RT). Biosynthesis of the msDNA starts with the transcription of all three genes (Fig. 2.6B). Depending on the retron element, all three genes are transcribed either into one mRNA or two molecules, one encompassing the *msr* and *msd* genes and the second the *ret* gene. The transcript from the *msr–msd* region is folded into a secondary structure using internal repeats. The RT utilizes the 2′ OH group of the branching G residue to initiate cDNA synthesis from the folded mRNA transcript that serves as both a primer and template. Upon completion of cDNA synthesis, part of the RNA not used as a template for cDNA synthesis remains attached to the cDNA, thus producing the satellite molecule, msDNA, recovered from the host cell.

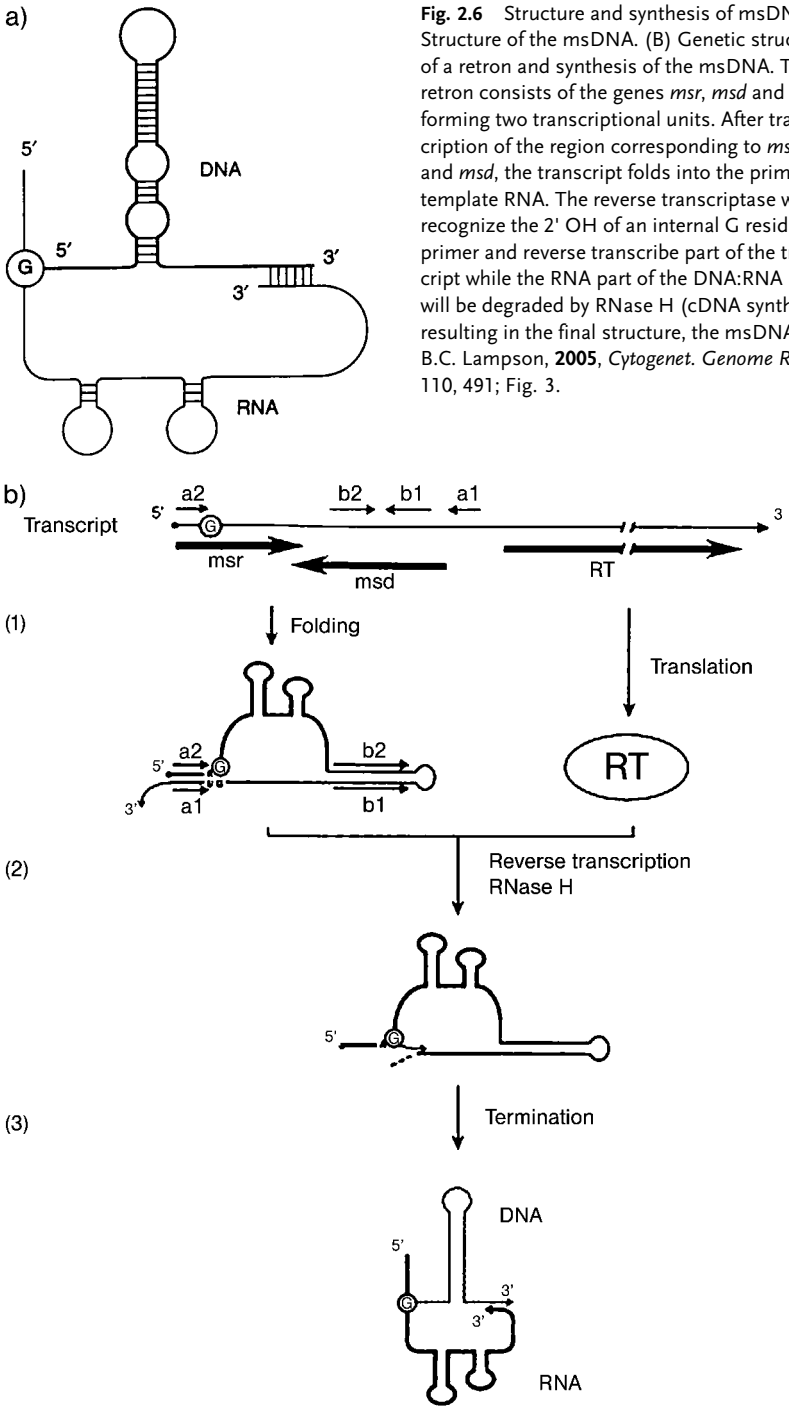


Fig. 2.6 Structure and synthesis of msDNA. (A) Structure of the msDNA. (B) Genetic structure of a retron and synthesis of the msDNA. The retron consists of the genes *msr*, *msd* and *ret*, forming two transcriptional units. After transcription of the region corresponding to *msr* and *msd*, the transcript folds into the primer-template RNA. The reverse transcriptase will recognize the 2' OH of an internal G residue as primer and reverse transcribe part of the transcript while the RNA part of the DNA:RNA hybrid will be degraded by RNase H (cDNA synthesis) resulting in the final structure, the msDNA. B.C. Lampson, **2005**, *Cytogenet. Genome Res.* 110, 491; Fig. 3.

Genetic elements that code for RT and employ the process of reverse transcription in some stage of their replication or mobility care called *retroelements*. Two different types of prokaryotic retroelements have been described:

- group II introns which may use reverse transcription to mobilize the intron element to new locations (see Section 6.3.5)
- retrons, the elements coding for msDNA.

Retrons have been described in several other bacterial species and families including pathogenic *E. coli*, *Salmonella*, *Proteus*, *Klebsiella*, *Vibrio* and *Rhizobium*. The function of msDNA is unknown, but it has been suggested, at least in some cases, that retrons may play a role in pathogenicity.

2.4 Large Rearrangements Within the Chromosome

One of the consequences of DNA replication and repair is the formation of rare chromosomal rearrangements, namely deletions, tandem duplications and inversions. Deletions can remove multiple functions and are irreversible. Duplications can amplify a coding region, are highly reversible and can be considered a temporary regulatory state. Inversions change the orientation of a sequence in the chromosome and often change the expression of one or more genes. Most chromosomal rearrangements involve either direct (duplication, deletion) or indirect repeated sequence elements. Bacterial chromosomes can be rearranged by two different mechanisms:

- homologous recombination between repetitive sequences
- sequence-specific recombination.

We will discuss a few examples. Duplications and further amplifications can result in the tandem duplication of antibiotic resistance genes, leading to an increase in resistance against the appropriate antibiotic. Chromosomes of Streptomycetes may contain amplifiable units allowing their duplication up to 500 times. Small deletions can alter the gene dosis, the gene product or the pattern of gene expression. A few examples will be discussed. Inversion of DNA sequences within the chromosome is used to turn gene expression on and off or to produce two different gene products (see Section 4.2.3).

2.4.1 Duplications

Duplication of chromosomal DNA is defined by copying the sequence either from one region of the DNA to another or, more commonly, a sequence is immediately followed by its duplicate (tandem duplication). Tandem duplications can result from recombination between directly repeated sequences in DNA, as shown in Fig. 2.7. Pairing between two directly repeated sequences in two different DNA

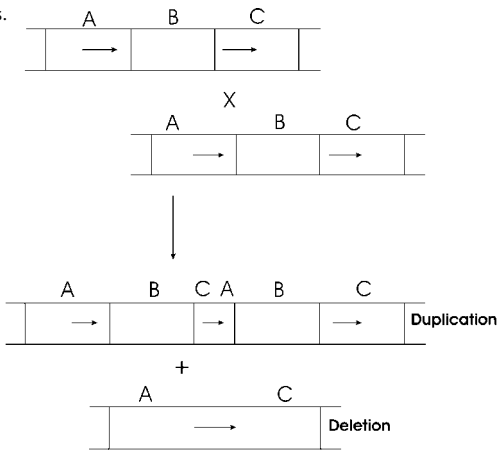


Fig. 2.7 Formation of tandem duplications. Direct repeats (shown as arrowheads) within genes A and C lead to pairing followed by recombination. The outcome of this reaction is a tandem duplication in the upper and a deletion in the lower chromosome.

molecules leads to a duplication in one and a deletion in the second DNA molecule.

Increase in Antibiotic Resistance by Gene Amplification

In some cases, selection for increased gene dosage results in amplification of short chromosomal segments up to 100 times. Gene amplification has been observed usually in the response of drug-resistance genes to high levels of antibiotic. One example is the *ampC* gene of *E. coli* K-12 coding for a β-lactamase which can hydrolyze the β-lactam ring of ampicillin. Ampicillin resistance is strictly related to the *ampC* gene copy number. The repeat size of *ampC* has been determined to be 9.8 kb, which is flanked by a 12-bp sequence of perfect homology; and a repeat copy number of up to 50 has been determined. Based on this and other known amplifications, recombination between randomly occurring short homologies (12–13 bp) could be a general mechanism to generate tandem duplications.

Tandem Duplication of Amplifiable Units in Streptomycetes

Species of *Streptomyces* can amplify certain regions of their chromosome. For example *S. fradiae* contains an amplifiable unit of 8.3 kb flanked by 2.2-kb direct repeats. This unit can be amplified up to 500 times, corresponding to about 5.2 Mbp!

2.4.2 Deletions

Deletions may be very long, removing many genes and only limited by essential genes whose deletion will lead to cell death. In *E. coli*, about 5% of all spontaneous mutations are deletions. Deletions can be either caused by recombination between

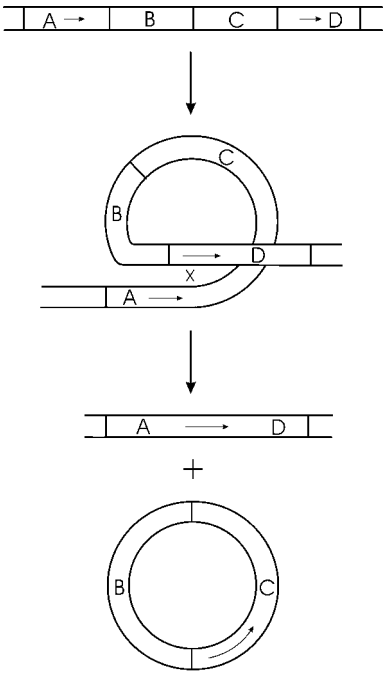


Fig. 2.8 Deletion between direct repeats located on one chromosome. Both repeats pair by looping out the DNA in between, followed by homologous recombination. Since the deleted circular DNA molecule does not contains an origin of replication, it is lost during subsequent cell divisions.

direct repeats within the same chromosome (Fig. 2.8) or between direct repeats involving two chromosomes (Fig. 2.7).

Turning Gene Expression On and Off by Small Deletions

N. gonorrhoeae is a Gram-negative bacterial pathogen highly adapted to survive within a single host population, humans. The cells colonize mucosal surfaces of the genito-urinary tract following sexual transmission of gonococci. Primary attachment to mucosal epithelia is mediated by pili and a more intimate association is then thought to be established by the colony *opa*city-associated (Opa) proteins. A single strain can possess up to 12 unlinked chromosomal alleles that encode distinct Opa variants. Their expression is phase-variable owing to RecA-independent DNA rearrangements that alter the number of pentanucleotide coding repeat (CR) units in the leader sequence and, thereby, influence the reading frame of these constitutively transcribed *opa* genes (Fig. 2.9). Phase variation of individual loci occurs at a frequency of ~10^{-3} *in vitro*, constantly generating a heterogenous population of bacteria expressing none, one or multiple Opa proteins.

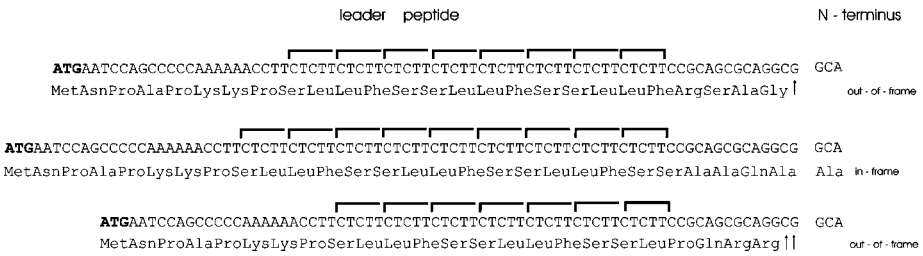


Fig. 2.9 Small deletions within the coding region of the signal peptide of the Opa protein can turn on and off the corresponding gene. The coding region for the leader peptide can contain a variable number of the pentanucleotide coding repeat units (indicated by brackets) which determine the reading frame.

Large Deletions Within the Chromosomes of *B. subtilis* and *Anabaena* Lead to Gene Expression

The ability to respond to environmental changes is an essential feature in the survival of microorganisms. Certain bacteria undergo development in response to changes in their environment, involving DNA rearrangements that are strictly under developmental control. DNA rearrangements have been reported in two bacterial species: *B. subtilis* and nitrogen-fixing cyanobacteria. Development in these organisms occurs in response to environmental cues such as nutrient deprivation. The *B. subtilis* and cyanobacterial DNA rearrangements show strict cell-type specificity.

DNA Rearrangement in *Anabaena*

Anabaena strain PCC 7120 is a filamentous cyanobacterium that is capable of both oxygenic photosynthesis and aerobic nitrogen fixation. Filaments contain only vegetative cells when grown in media containing a source of combined nitrogen, but under nitrogen-limiting conditions approximately one in every ten vegetative cells differentiates into a *heterocyst*. While vegetative cells carry out O_2-evolving photosynthetic reactions, heterocysts are terminally differentiated cells that act as anaerobic factories for fixing atmospheric nitrogen. The ability to fix atmospheric nitrogen depends on the oxygen-sensitive enzyme nitrogenase. The differentiation process begins shortly after nitrogen deprivation and requires 20–24 h to complete. Heterocysts supply fixed nitrogen as glutamine and other amino acids to the neighboring vegetative cells, which supply heterocysts with fixed carbon produced by photosynthesis.

Heterocyst development involves programmed deletion of two DNA elements, the 11-kb and 55-kb elements (Fig. 2.10). The 11-kb element is located within the open reading frame of the *nifD* gene and is bordered by 11-bp direct repeats; and site-specific recombination within these direct repeats results in deletion of the element from the chromosome, catalyzed by the *xisA* gene (coding for the XisA recombinase) located within the excised element (Fig. 2.10). The excised 11-kb element is found as a stable circular molecule in heterocysts. Excision of the 11-kb

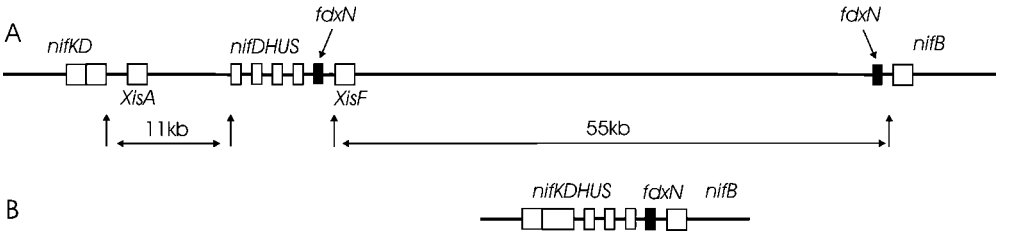


Fig. 2.10 Deletion of two different DNA segments during heterocyst differentiation in *Anabaena*. (A) In vegetative cells, an 11-kb element splits the *nifD* gene and a 55-kb element separates *nifB* from its promoter. (B) During heterocyst development, the 11-kb and 55-kb segments of DNA are excised from the chromosome by site-specific recombination. Removal of the 11-kb segment fuses the two parts of the *nifD* gene, while deletion of the 55-kb segment brings the *nifB* gene close to the *nifKDHUS* operon.

element exerts two significant effects on *nif* gene expression. First, it fuses the 5′ and the 3′ of the *nifD* open reading frame. Second, it places the *nifK* gene under the transcriptional control of the *nifH* promoter. The second genomic rearrangement has DNA recombination sites near the *nifS* and *nifB* genes and removes a 55-kb element flanked by direct repeats designated *fdxN* and catalyzed by the XisF recombinase. This deletion results in placement of the *rbcLS* operon about 10 kb from the *nifS* gene on the chromosome.

DNA Rearrangement in *B. subtilis*

The structural gene (*sigK*) for the mother cell RNA polymerase sigma factor σ^K in *B. subtilis* is a composite of two truncated genes, named *spoIVCB* and *spoIIIC*, which are brought together by site-specific recombination during sporulation (Fig. 2.11). This recombination event is compartmentalized in that the mother cell, but not the forespore chromosome, is governed by the product of a gene named *spoIVCA* located in the excised DNA. Excision of the 48-kb element termed *skin* (*s*igma *K in*tervening) joins *spoIIIC* (encoding the carboxy-terminal portion of σ^K) and *spoIVCB* (encoding the amino-terminal portion). The 55-kDa SpoIVCA protein binds specifically to 21-bp indirect repeats flanking the *skin* recombination sites (5-bp direct repeats) and releases the intervening DNA from the chromosome as a circle.

Deletions at the Ends of the Linear *Streptomyces* Chromosomes

The linear *S. coelicolor* M145 chromosome consists of a central portion of 8.7 Mb encoding genes likely to be essential for growth, while the chromosome arms contain many genes with nonessential functions, such as those of hydrolytic and secondary metabolism enzymes. The ends of *Streptomyces* chromosomes are known to be highly dynamic, varying greatly in structure. The inverted repeat sequences identified at the chromosome ends (termed *t*erminal *i*nverted *r*epeats; TIRs) range

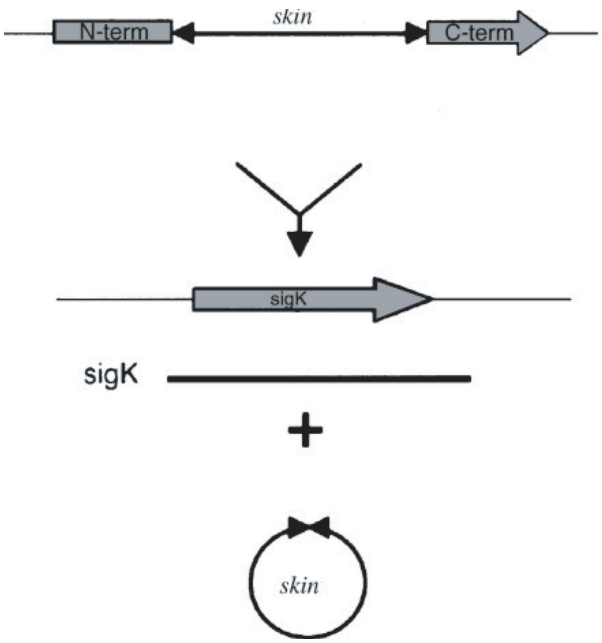


Fig. 2.11 Deletion of the 48-kb *skin* element restores an active *sigK* gene. Activation of the site-specific SpoIVCA recombinase within the mother cell compartment of sporulating *B. subtilis* cells leads to the excision of the 48-kb element joining the two halves of the *sigK* gene. J.D. Haraldsen, A.L. Sonensheim **2003**, *Mol. Microbiol.* 48, 811; Fig. 2, modified.

from 175 bp for *S. avermitilis* to 850 kb for a *S. ambofaciens* mutant; and the sequence of only the first ~200 bp of TIRs appears to be conserved among *Streptomyces* species. The ends of *Streptomyces* chromosomes undergo large deletions at frequencies between 0.1% and 1.0%, removing 300–800 kb, and these deletions are often accompanied by circularization of the chromosome and amplification of subtelomeric DNA. In addition, *Streptomycetes* can exchange sequences between chromosome ends and between the ends and a linear plasmid. Transposable elements in the arms of Streptomycetes chromosomes may participate in frequent genome arrangements. It has been shown that 40% of the 109 transposase homologs in *S. coelicor* M145 are located within 100 kb of each chromosome end. These transposable elements have the potential to mediate DNA insertion, deletion, inversion and exchange events through either site-specific transposition or homologous recombination.

2.4.3 Inversions

Inversions of DNA sequences are caused in the same way as deletions, but this time involving indirect repeats (Fig. 2.9). In contrast to deletions, inversion can re-

vert. This is used to turn genes on and off (either the promoter or the gene itself is located within the invertible region) or to create two different gene products. In the latter case, the promoter and the 5′ region of the gene is located outside of the invertible region whereas the two different 3′ regions are situated within the invertible segment. Examples are the G-segment of phage Mu and the C-segment of phage P1, where the two variable proteins S and S′ constitute the tail fibers involved in attachment to host cells (see Section 4.2.3). The constant N-terminal part of the two proteins is assumed to bind to the phage tail, whereas the variable C-terminal part attaches to the host cells. Inversion systems within the chromosome influence expression of pili (*E. coli fim* genes) or flagella (*S. typhimurium flj* genes; see Section 4.2.3).

3
The Bacterial Cell Cycle: Replication of the Chromosome, Partitioning and Cell Division

Under growth conditions, bacteria duplicate their genome, partition it towards the two poles (studied in rod-shaped bacteria only) and finally divide at the correct site in the cell at the correct time with high precision. Since replication and cell division has been studied most thoroughly in *E. coli*, this enteric bacterium will serve as the paradigm organism.

Replication of the circular *E. coli* chromosome is initiated at the orgin of replication, *oriC*, and terminates in the termination region, *terC*, located diagramatically opposite *oriC*. It involves 25–30 proteins, most of which are essential. After termination of replication, the two daughter chromosomes are often interwound at the termination region, forming *catenanes* which are resolved by a topoisomerase. Next, the two chromosomes have to be faithfully segregated (partitioned) into the two daughter cells, a process not yet understood very well. It involved proteins for compaction of the chromosomes as well as proteins separating them. The last step is cell division and requires ten identified proteins forming the division complex. In addition, three proteins assure formation of the septum at midcell and prevent its formation near the two poles.

3.1
Replication

As outlined under Section 2.1, most bacterial species carry circular chromosomes. Therefore and because of our knowledge accumulated over the past decades, we will consider replication of the circular *E. coli* chromosome in detail. At the end, we will review our current limited understanding of how linear chromosomes are replicated.

3.1.1
Replication of Circular Chromosomes

Replication of the circular chromosome starts at a unique site, the replication origin *oriC*, and proceeds bidirectionally from there to the terminus region, *terC*. Chromosome replication is a highly complex and regulated process that involves

Dynamics of the Bacterial Chromosome. Structure and Function. W. Schumann
Copyright © 2006 Wiley-VCH Verlag GmbH & Co. KGaA
ISBN 3-527-30496-7

25–30 proteins assembled together to form a multienzyme machine designated as the *replisome*. Replication origins of different bacteria have varying sizes, but all contain several binding sites for an initiator protein termed *DnaA*. The DnaA protein is absolutely necessary for opening the DNA duplex in the *oriC* region, which is then kept single-stranded by the *s*ingle-stranded DNA-*b*inding (SSB) protein. Next, two *replicative helicase DnaB* complexes are loaded onto the two single-stranded DNA strands by the *DnaC* loader which recruit the *DNA primase protein DnaG* and separate the two complementary strands until the *terC* region is reached, where they are stopped by the *contrahelicase protein Tus*. The DnaG protein synthesizes *RNA primers*, only two on the two *leading strands* and 2000–3000 on the two *lagging strands* which are elongated by the *DNA polymerase III* (Pol III), consisting of four components: the *processivity factor* (β clamp), the *catalytic component* (Pol III core), the *clamp loader complex* and the *dimer stabilization complex*. The RNA primers at the lagging strands are elongated by Pol III to yield *Okazaki fragments*. In a last step, most RNA primers are removed primarily by the *RNase H* and any remaining ribonucleotide is hydrolyzed by the 5′→3′ exonuclease activity of *DNA polymerase I* (Pol I) which also fills the gap, using its polymerizing activity. If the Pol I has arrived at the first nucleotide of the downstream Okazaki fragment, it is unable to join the 3′ end of the last added nucleotide to the 5′ end of this nucleotide; and this last step is carried out by the *DNA ligase* to seal these nicks. When replication of the mother chromosome has finished, the two daughter chromosomes might not be completely separated. Instead, they might either form a *dimer* or be *catenated*. While a *site-specific resolution system* will resolve a dimer into the two monomers, *topoisomerase IV* will catalyze *decatenation* of the two interlinked chromosomes. Control of replication occurs at the initiation stage and involves three mechanisms:

1. *sequestration*: transient binding of the hemimethylated *oriC* to the membrane by the *SeqA* and *SeqB* proteins;
2. *titration of DnaA*: DnaA proteins are stored to specific sites on the chromosome;
3. *inactivation of DnaA-ATP*: the proteins *IdaB* and *Hda* mediate the interaction between ATP-DnaA and the β-subunit sliding clamp of the DNA polymerase III, which leads to ATP hydrolysis.

3.1.1.1 Components of the *E. coli* Replisome

In bacteria, replication is initiated at one specific DNA sequence, the *oriC*, during the cell cycle and proceeds with circular chromosomes bidirectionally to another specific DNA sequence, the *terC*. The *E. coli oriC* is defined as a 245-bp minimal DNA sequence, based on mutational analysis of this sequence. 25–30 proteins are assembled at the *oriC* and form a multienzyme complex termed the *replisome*. One replisome is formed per DNA strand which is then converted into the *replication fork* when the replisome starts to move towards the *terC* region. Once assembled, each highly processive replisome proceeds bidirectionally at more than

50 Kbp min^{-1} from *oriC* to *terC* for termination, while making only one misincorporation mistake for every 10^{10} nucleotides polymerized. This means that approximately one in every 2000 progeny bacteria will have a point mutation in its DNA. The process of replication has been divided into the three distinct stages of initiation, elongation and termination. Before we deal with these three stages in detail, we will describe the known proteins involved in these different stages; these are summarized in Table 3.1.

Assembly of the replisome and initiation of replication starts within the supercoiled *oriC* region, which is defined as a 260-bp DNA sequence containing binding sites for several proteins as well as important signatures presented in Fig. 3.1. Of central importance are the five DnaA binding sites R1 through R4 and M called *DnaA boxes*. Furthermore, *oriC* contains binding sites for accessory proteins like IHF and FIS and the control factors IciA, Rob and HNS, an AT-rich region

Table 3.1 Proteins and their function involved in replication of the *E. coli* chromosome.

Gene	*Function*
dam	DNA methyltransferase; methylates A residues in the GATC sequence
dnaA	Initiation of replication; DNA-binding protein (DnaA boxes); active form ATP-DnaA; opens the DNA double helix at *oriC*
dnaB	Replicative helicase; homohexamer
dnaC	Helicase loader; loads the helicase onto single-stranded DNA; homohexamer
dnaG	DNA primase; synthesizes RNA primers (10±1 nucleotides), one on each leading strand and 2000–3000 on the two lagging strands
fis	Binds to *oriC*; involved in structuring of the *oriC* region
gyrAB	DNA gyrase; heterodimer αβ; binds to specific sites and removes positive supercoils
hda	Mediates interaction between ATP-DnaA and the β-subunit sliding clamp
himAD	Two subunits of IHF; heterodimer
hns	Binds to *oriC* and enhances unwinding at a high ATP concentration (>2 mM)
hobN	Involved in the formation of the membrane complex consisting of hemimethylated *oriC*, SeqA and SeqB
idaB	Stimulates ATP hydrolysis of ATP-DnaA
iciA	Binds to the AT-rich 13mers; prevents strand opening at *oriC*
ligA	NAD-dependent DNA ligase; seals DNA nicks
rnhA	RNase H; removes the RNA primers
rob	Binds to the *oriC*; function largely unknown
polA	DNA polymerase I; removes RNA primers and fills the gaps
seqA	Involved in sequestration of the hemimethylated *oriC* to the inner membrane
seqB	Enhances the ability of SeqA to tether the hemimethylated *oriC* to the inner membrane
ssb	Single-stranded DNA binding protein; prevent renaturation of the two separated strands
tus	Tus proteins binds to 22-bp *ter* sequences and acts as a contrahelicase

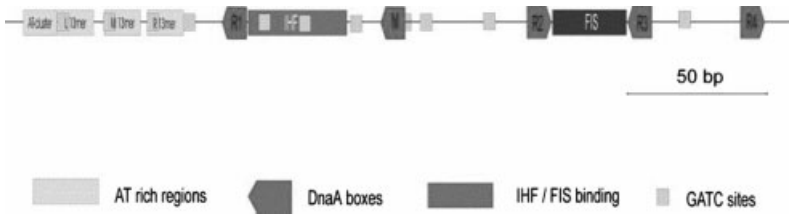


Fig. 3.1 Organization of the *oriC* region of *E. coli*. The *oriC* region of *E. coli* consists of five DnaA boxes called R1 through R4 and M, binding sites for the histone-like proteins FIS and IHF, three AT-rich regions and 11 GATC methylation sites. W. Messer **2002**, *FEMS Microbiol. Rev.* 26, 355; Fig. 1. (This figure also appears with the color plates.)

consisting of three 13-mer repeats, the so-called AT-cluster. In addition, 11 GATC sites, recognition sequences for the Dam methyltransferase, play an important role in the regulation of replication initiation (see below).

As already mentioned, the replication mechanism involves between 25–30 different proteins (see Table 3.1) where some are needed only during the initiation process, one just for termination, but most are needed for initiation and elongation. Most of these proteins have been identified by temperature-sensitive mutations in their encoding genes. A closer analysis of these conditional lethal mutations led to their classification as *fast-* and *slow-stop* mutants. Fast-stop mutants stop replication immediately when growing cells are shifted to the restrictive temperature, while the slow-stop mutants take about 60 min, indicating that, in the first case, elongation is involved and initiation in the latter.

DnaA (51 kDa) is the key protein for the initiation of chromosome replication in *E. coli* and other bacteria. All functions of the active DnaA protein, ATP-DnaA, depend on its ability to interact specifically with two different asymmetric DNA sequences, a 9-bp recognition sequence (5′-TTATNCACA-3′) and a 6-mer sequence (5′-AGATCT-3′ or a close match, present within the AT-rich region of the *oriC*), denoted as DnaA boxes. Besides stimulating the initiation of replication, the DnaA protein can also act as a transcription factor. DnaA boxes are found in the promoter regions of many genes, where they can mediate repression, transcriptional activation, or transcriptional termination due to loop formation between two DnaA boxes in a transcriptional unit and long-range DnaA–DnaA interaction. A threshold level of ATP-DnaA protein is required for the conversion of the initial to the open complex. Fis protein has a negative effect on the reaction, while HU and IHF proteins, at a high ATP concentration (>2 mM), enhance unwinding. The replicative active complex contains 20–30 DnaA monomers. An average *E. coli* cell contains 1000–2000 DnaA molecules and this level does not vary appreciably during the cell cycle. DnaA is also involved in the initiation of replication of many *E. coli* plasmids. These carry in their replication origins, termed *oriV*, one or several DnaA boxes and an AT-rich region for unwinding, and additionally, several *iterons*, repeated binding sites for a plasmid-encoded initiation protein. In the case of plasmids, DnaA works in concert with the plasmid-encoded

initiation protein. In the case of the F factor, the *oriV* consists of two DnaA boxes, followed by an AT-rich region and four 19-bp direct repeats serving as iterons for the plasmid-encoded initiator protein RepA. Efficient unwinding needs the concerted action of DnaA, RepE and HU.

Proteins *DnaB* (52 kDa) and *DnaC* (27 kDa) form a complex in solution, where both proteins exist as homohexamers; and this complex $DnaB_6$–$DnaC_6$ is stabilized by ATP bound by DnaC. The major function of DnaC is to escort DnaB to the *oriC*. ATP stimulates ssDNA binding by DnaC, leading to expansion of the ssDNA bubble at the origin. Two cofactors, ssDNA and DnaB, trigger the hydrolysis of ATP, converting DnaC to the ADP form that no longer inhibits DnaB. This finding led to the idea that DnaC functions as a dual ATP/ADP switch protein, where both the ATP and the ADP forms are sequentially required for replication. The *dnaB* gene codes for a replicative helicase, a motor protein that unwinds the DNA. DNA helicases are the motor proteins responsible for separating individual strands of dsDNA to provide ssDNA for key cellular processes such as DNA repair, recombination and replication. DnaB is a member of the hexameric DNA helicase family which show a common polarity movement with 5′→3′ polarity. DnaB is a true multifunctional enzyme with a number of distinct enzymatic activities, including ATP hydrolysis, DNA binding and association with other replicative proteins, such as DnaG primase and Pol III holoenzyme. The interactions with replication proteins in the replication fork allow it to form the replisome with primase and holoenzyme dimer in order to replicate the leading and lagging strands simultaneously. The energy of DNA-dependent ATP hydrolysis allows duplex DNA unwinding by the DnaB helicase and movement of the replisome synchronously with the progression of DNA unwinding and fork movement. Two DnaB hexameric helicases are assembled onto the ssDNA bubble, one on each strand, to form bi-directional replication forks. The DnaC helicase loader is required in this process, but dissociates from DnaB after loading it onto DNA. Each of these proteins binds and/or hydrolyzes ATP.

Since all known DNA polymerases are incapable of *de novo* initiation, cells use other mechanisms to prime DNA synthesis throughout replication. In 1971, A. Kornberg proposed that DNA replication initiation required RNA transcription. Since then, replication-priming RNA polymerases (*primases*) have proven central to cellular replication mechanisms. Primases initiate leading-strand synthesis once and lagging-strand synthesis multiple times during the course of replication. The *E. coli DnaG primase* is active as a 65.6-kDa monomer and interacts with the replicative DnaB helicase, *s*sDNA-*b*inding *p*rotein (SSB) and χ, a subunit of the clamp-loading γ-complex of Pol III (see below). Two primers are synthesized on the two leadings strands, and on the lagging strands, a new primer is synthesized about once very 2 kb (2000–3000 RNA primers in total), recognizing the sequence 3′-GTC-5′ and starting synthesis opposite the T residue. Each primer is 10–12 nucleotides long. Removal of the primers occurs presumably by the combined action of *RHase H1* and the 5′→3′ exonuclease activity of *Pol I*.

The *ssDNA-binding protein SSB* is a tetramer of an 18.8-kDa polypeptide. SSB binds single-stranded DNA tightly and cooperatively; and the SSB-coated DNA ap-

pears in the electron microscope as a ''beads on a string'' type structure. It has been suggested that this arises from the wrapping of 145 nucleotides of ssDNA around the SSB tetramer. For SSB, three properties related to DNA replication have been described:

1. SSB is assumed to protect significant stretches of ssDNA, at least equal to the size of the most current nascent Okazaki fragment.
2. SSB can stimulate both the rate and the processivity of the Pol III by melting regions of secondary structure in the DNA that are normally inhibitory to the passage of the polymerase.
3. There may be specific interactions between SSB and other enzymes at the replication fork. For example, the χ subunit can interact with SSB.

DNA polymerases (Pol) catalyze the addition of a nucleotide onto an existing 3′ OH of a growing DNA chain. They synthesize short stretches of DNA, but tend to fall off the template DNA and exhibit a distributive rather than a processive behavior. To function in a more processive manner during the replication of large genomes, DNA polymerase accessory proteins tether the polymerase to the DNA template. The most commonly used mechanism is a sliding clamp that is topologically linked with the double-stranded DNA and binds to the polymerase. These clamps form a ring-shaped complex made up of dimers in the case of the *E. coli* Pol III. How does the strand of circular chromosomal DNA get inside the ring? This function is carried out by the clamp loading machine, the γ complex in *E. coli*.

All known polymerases synthesize nucleic acid in a 5′→3′ direction. This feature requires that the two antiparallel strands of DNA be replaced asymmetrically. While the *leading strand* is made continously at the replication fork, the *lagging strand* is formed discontinuously. Because DNA polymerases are incapable of *de novo* initiation, cells use other mechanisms to prime DNA synthesis throughout replication. In 1971, Kornberg and co-workers proposed that DNA replication initiation requires RNA transcription. Since then, replication-priming RNA polymerases (*primases*) have proven central to cellular and many viral replication mechanisms. Primases initiate leading-strand synthesis once and lagging-strand synthesis multiple times during the course of replication. Depending on the organism, primases exist either as individual proteins or as primase–helicase polyproteins; and in almost all cases their activities are coupled to the replisome by protein–protein interactions with other replication factors. *E. coli* primase (DnaG) interacts with the replicative DnaB helicase, SSB and Pol III holoenzyme. Although DnaG is capable of synthesizing 60-nucleotide primers *in vitro*, this primer length is restrained to 10(±1) nucleotides in the context of the replisome. During lagging-strand synthesis of the *E. coli* genome, DnaG proteins must transcribe 2000–3000 RNA primers at a rate of about 1 primer s^{-1}. In *E. coli*, primase initiates primer RNA synthesis preferentially at the T residue of a 5′-CTG-3′ template sequence.

DNA gyrase (see Section 2.1.1) is required to remove the supercoils. In the 4.6-Mbp *E. coli* chromosome, some 4.6×10^5 links must be removed during each

round of replication at a rate of ~10^2 links s^{-1}. Failure to remove even a single link will result in the sister chromosomes being catenated and prevent sister chromosome separation. The DNA gyrase acts ahead of the replication fork to remove positive supercoils as they accumulate ahead of the fork.

DNA ligases catalyze the sealing of 5′ phosphate and 3′ hydroxyl termini at nicks in duplex DNA by means of three sequential nucleotidyl transfer reactions. In the first step, attack on the adenylyl α phosphorus of NAD^+ by ligase results in release of nicotinamide mononucleotide and formation of a covalent intermediate (ligase-adenylate) in which AMP is linked through a phosphoamide bond to lysine. In the second step, the AMP is transferred to the 5′ phosphate at the nick to form DNA-adenylate. In the third step, the unadenylated ligase catalyzes attack by the 3′ OH of the nick on DNA-adenylate to join the two polynucleotides and release AMP. One NAD^+-dependent DNA ligase is found in every bacterial species. NAD^+-dependent ligases are of fairly uniform size (656–837 amino acids) and extensive amino acid sequence conservation occurs throughout the entire length of the polypeptides. Postgenomic analysis and biochemical studies have revealed that many bacteria contain one or more ATP-dependent DNA ligases in addition to the NAD^+-dependent enzyme. Therefore, bacteria may exploit different DNA ligase isoenzymes for different physiological functions (e.g., replication, repair, homologous recombination, nonhomologous endjoining). *E. coli* contains a second NAD^+-dependent DNA ligase encoded by gene *ligB*, the function of which remains elusive.

A total of five DNA polymerases (Pol I through Pol V) have been discovered in *E. coli* up to now, where Pol III is primarily responsible for the efficient and accurate replication of the 4.6-Mbp *E. coli* chromosome, many phage DNAs and plasmids. Pol III holoenzyme performs this task with high speed and accuracy with the help of ten component subunits organized in four types of functional complexes as follows: (1) the sliding clamp processivity factor (β_2 clamp), (2) the energy-dependent clamp loader ($\gamma_3\delta\delta'\chi\psi$), (3) the Pol core ($\alpha$-$\varepsilon$-$\theta$) and (4) the dimer stabilization complex (τ, γ). All the genes and their functions are summarized in Table 3.2.

The β *clamp* is a ring with a central cavity of about 35 Å in diameter that is sufficiently large to encircle duplex DNA without steric hindrance and is formed by two crescent-shaped protomers. The β clamp slides on the single-stranded DNA and tethers the catalytic subunits of the Pol (the core complex) to the template. It confers upon the polymerase the ability to faithfully track the rapidly moving replication fork while synthesizing leading and lagging strand DNA simultaneously. The closed circular β clamp must be opened frequently at one interface during DNA replication for assembly on DNA to initiate processive replication as well as for disassembly of the β ring from DNA when replication is complete. This reaction is carried out by the *clamp loading complex*, also called the γ complex, consisting of five different proteins forming the $\gamma\delta\delta'\chi\psi$ complex. The γ subunit powers the complex by binding and hydrolyzing ATP, δ′ regulates the activity of the complex by binding δ, and when the complex is in an open form, δ functions to bind to and destabilize the β subunit. The χ and ψ subunits are not essential, but aug-

Table 3.2 Subunits of Pol III and their function.

Gene	*Protein (kDa)*	*Function*
Sliding clamp		
dnaN	β (40.6)	Homodimer; sliding clamp
Clamp loader		
holA	δ (38.7)	Binds and destablizes the β-subunit in the open form
holB	δ′ (40.0)	Regulates the activity of the clamp loader
dnaX	γ (47.0)	Binds clamp loader and SSB protein
holC	χ (16.6)	Recognizes primer/template through interaction with SSB
holD	ψ (15.0)	Recognizes primer/template through interaction with SSB
Pol III core		
dnaE	α (13.0)	Polymerase activity
dnaQ	ε (27.0)	3′→5′ proofreading activity
holE	θ (8.3)	Stabilizes the proofreading function
Dimer stabilization complex		
dnaX	τ (71.0)	Joins leading and lagging Pol III

ment clamp loading activity by facilitating primer/template recognition through interactions with SSB.

The Pol III holoenzyme contains two molecules of *polymerase core* for simultaneous synthesis of both strands of the DNA helix: subunit α (conferring the polymerase activity), ε (the proofreading 3′→5′ exonuclease), and θ (protein stabilizing ε). The subunits are arranged in a linear fashion, α-ε-θ, with ε binding both α and θ. The ε subunit consists of two domains, an N-terminal domain containing the exonuclease and a C-teminal domain essential for binding the α subunit. Rapid and processive DNA synthesis of Pol III holoenzyme is dependent on the interaction between the α subunit of the core polymerase and the β clamp. Once attached to the β subunit, the catalytic α subunit of the polymerase can move along DNA for tens of kilobases without dissociation (high processivity), incorporating new nucleotides into the growing DNA strand at speeds as high as 750 nt s^{-1}.

Two copies of Pol III, one on the leading strand and one on the lagging strand, are linked together by the *dimer stabilization complex*. This complex consists of the two τ subunits. The γ and τ subunits are encoded by the same *dnaX* gene. τ is the full-length product, and γ is truncated by a –1 ribosomal frameshift (see Section 6.4.2). The 24-kDa C-terminal sequence of τ binds both the core and DnaB helicase. The lagging strand is synthesized as short, 1–3 kb, Okazaki fragments. The DNA primase initiates each fragment by synthesis of a short RNA primer. As the Pol III extends the primer, it synthesizes DNA in the direction opposite the fork movement even though it travels with the fork due to its connection to both DnaB (through the τ subunit) and the leading Pol III (through the shared clamp loader).

These opposite motions are thought to be resolved by the formation of a DNA loop. As the DNA is extended, the loop grows until the Pol III arrives at the next RNA primer. At this point, the Pol III must dissociate from the DNA to start extension of the next Okazaki fragment. Given the very low copy number of Pol III holoenzyme (10–20 copies per cell) and the need for 2000–4000 Okazaki fragments, this lagging-strand Pol III must be rapidly recycled. It has been shown that the lagging-strand Pol III, upon completing a fragment, dissociates from the β clamp and binds to a new clamp, a mechanism called "*processivity switch*".

While the *E. coli* chromosome codes for only one DNA polymerase essential for its duplication, the fully sequenced genomes of the *Bacillus/Clostridium* group and the genome of the thermophilic microorganism *T. maritima* code for two genes with catalytic DNA polymerase. In the case of *B. subtilis*, it has been shown that there are two different essential DNA polymerases at the replication fork. While $dnaA_{BS}$ appears to be involved in the synthesis of the lagging DNA strand, *polC* is involved in the leading strand synthesis.

DNA polymerase I (102 kDa), the first polymerase discovered, is the most abundant polymerase within the cell (about 400 molecules per cell) and functions primarily to fill DNA gaps that arise during DNA repair, recombination and replication. This polymerase has two functional domains located on the same polypeptide: a 5′→3′ polymerase (near the C-terminus) and a 5′→3′ exonuclease (near the N-terminus). *E. coli* Pol I contains in addition a proofreading 3′→5′ exonuclease domain which is able to excise nonpairing NMPs.

In the *E. coli* chromosome, the adenine residues of GATC sequences are recognized and methylated by the *Dam methyltransferase* (for *D*NA *a*denine *m*ethylation). This protein transfers a methyl group from the methyl donor, *S*-adenosylmethionine, to the 6-amino group of adenine residues in hemimethlylated and unmethylated GATC sequences. If the replication fork passes a GATC sequence, this sequence occurs transiently in a hemimethylated form where the parental strand still carries the methylated A, while the newly synthesized daughter strand does not. Conversion of DNA from the hemi- to the fully methylated state is a relatively rapid process. However, at *oriC*, methylation is inhibited for one-third of the cell cycle. This delay is due to sequestration of the hemimethylated *oriC* to the inner membrane through interaction with the 21-kDa integral membrane protein *SeqA* (from *seq*uestration *A*) which forms a sequestration complex together with SeqB (see below for more details). Dam methylase is a monomer with a molecular weight of 32 kDa; and each *E. coli* cell possesses approximately 130 molecules of Dam methylase. Methylation of GATC sites by Dam methylase is implicated in regulation of *E. coli* chromosomal replication (see below) and methyl-directed mismatch repair (see Section 5.2.1). Also, Dam function plays a role in regulation of certain genes, such as *dnaA*.

In addition, there is a group of accessory proteins, the roles of which are poorly understood. These proteins include the histon-like DNA-binding proteins *Fis* (*f*actor for *i*nversion *s*timulation) and *IHF* (*i*ntegration *h*ost *f*actor) for which binding sites have been identified within the *oriC* region (see Fig. 3.1). By binding to *oriC*, these proteins are thought to alter the topological structure, thereby aiding DnaA

to open the DNA duplex. Two other proteins, *Rob* (*r*ight *oriC b*inding) and *IciA* (*i*nhibitor of *c*hromosome *i*nitiation *A*; binds to the 13-mers) stimulate and inhibit initiation of replication, respectively.

3.1.1.2 Initiation of Replication

The initiation of replication can be divided into 12 different, successive steps (Fig. 3.2):

1. ATP-DnaA binds first to box R1.
2. This initial binding serves as an anchor for the cooperative binding of ATP-DnaA to three additional DnaA boxes. The bound ATP-DnaA molecules introduce a bend of 40° at each DnaA box.
3. Fis protein binds to its specific site; and this complex is called the "*initial complex*".
4. Fis leaves the complex and ATP-DnaA binds to the fifth DnaA box, R3.
5. IHF binds to its site and introduces a strong bend; and the AT-rich region becomes partially unwound (about 28 bp) leading to the "*open complex*".
6. ATP-DnaA binds to additional sites, 6-mer single-stranded DnaA boxes with the sequence 5′-AGATCT-3′, or a close match, located predominantly in the AT-rich region. This is a high-affinity interaction and cooperative binding to the double-stranded AT-rich region is therefore presumably the limiting step in the initiation reaction, followed by unwinding.
7. SSB extends the initial 28 bp of DNA to 44–52 bp.
8. Two double hexamers of the replicative helicase DnaB and the helicase loader protein DnaC, one for each replication direction, must be loaded into the unwound region. Since single-stranded DNA covered with SSB is not a substrate for DnaB, helicase must be positioned by DnaA. Two $DnaB_6$–$DnaC_6$ complexes are actually introduced into the bubble, thereby expanding it to about 65 bp, and the first DnaB hexamer is positioned onto the lower strand. The DnaB hexamer forms a ring around the single-stranded DNA and the channel through the helicase ring corresponds to about 20 nucleotides.
9. $DnaC_6$ dissociates from the complexes with concomitant ATP hydrolysis, resulting in two head-to-head helicase hexamers. This activates DnaB as a helicase and the two helicase complexes move past each other in a 5′→3′ direction.
10. Now the DNA primase can enter the complex and synthesize two leading strand primers.
11. The sliding clamp, a ring-shaped dimer of the β-subunit of DNA polymerase III, is loaded onto each primed template by the clamp loader, the polymerase III γ complex, followed by binding of two Pol III core dimers.
12. This activates the intrinsic ATPase activity of DnaA, in cooperation with a protein IdaB.

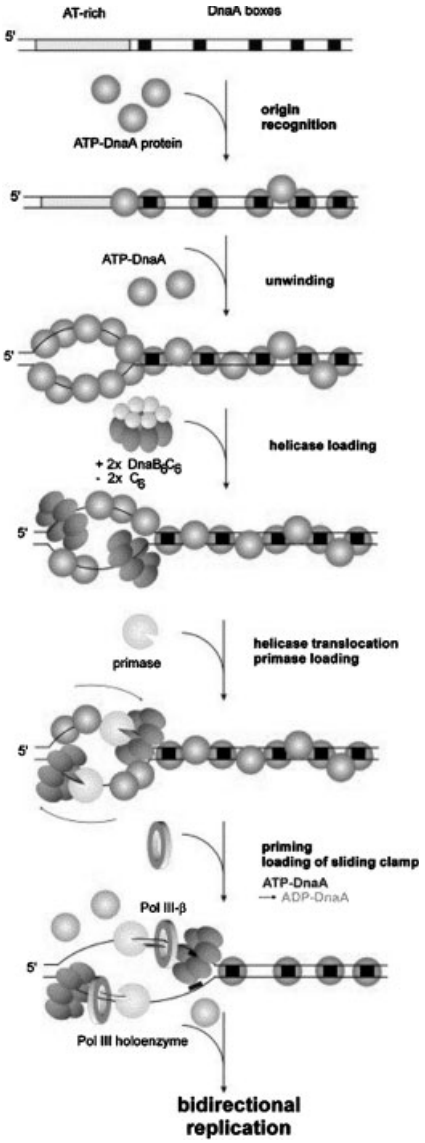


Fig. 3.2 The replication initiation cycle of *E. coli*. DnaA-ATP binds to the five DnaA boxes, which leads to the strand separation within the AT-rich region. Additional DnaA-ATP molecules bind the the ssDNA region to keep the strands apart. Next, DnaC loads one DnaB hexamer on each ssDNA strand, followed by dissociation of DnaC from the complex. DnaG primase binds to the DnaB replicative helicase and synthesizes the primer to which the β-sliding clamp of Pol III is loaded followed by the core complex. Now, bidirectional replication is started. W. Messer **2002**, *FEMS Microbiol. Rev.* 26, 355; Fig. 2.

The initiation of replication is ended when the Pol III starts elongating the primers. Where are the replication forks? It has been shown in *B. subtilis* that the two replication forks stay together at midpoint of the cell until replication is complete and the two *oriC*s move towards the poles of the cell.

3.1.1.3 Elongation of Replication

Replication forks usually proceed at a constant speed along the DNA (about 1000 bp s^{-1}). However, replication forks can also be stalled transiently at replication fork pausing sites or irreversibly at replication fork barriers. DNA downstream of a replication fork barrier has to be replicated by a converging fork that ultimately merges with the arrested fork.

During elongation, the replication forks often pause, stall and even collapse, causing double-strand breaks (DSBs) during their passage from the origin to the terminus. Such interruptions are postulated to be caused by single-strand breaks (SSBs) in the template DNA, either pre-existing or arising from abnormal DNA structures (folded DNA, cruciform structures, cross-links, etc.) or from collision with static or converging protein complexes. The cells must have developed highly effective strategies for dealing with replication fork arrests and/or strand breakage during replication without invoking the error-prone, mutagenic repair pathways (usually activated in response to strong DNA damage), such that the integrity and/or the organization of the bacterial genome could be maintained during chromosome replication through generations. Stalling of the replicative apparatus presents two major problems for all organisms. First, the original block to fork progression must be removed and any damage to the DNA must be repaired. Second, replication must be restarted. The first of these challenges is met by an array of repair systems which act to remove or bypass lesions (described under Section 4.1.4). The second challenge, that of restart replication, must be met by reloading the replicative machinery onto the DNA away from the normal preprogrammed origins of replication; and formation of a primosome is a prerequisite, requiring several new proteins.

Formation of a Replication Restart Primosome Replication restart does not require DnaA but employs PriA, an enzyme that recognizes specific branched DNA structures in which a leading strand is present at the branch point. At such structures, PriA performs the same function as DnaA at *oriC* by allowing to load DnaB onto the lagging strand template. Primosome assembly occurs in discrete steps at a *PAS* (*p*rimosome *a*ssembly *s*ite) sequence. PriA recognizes and binds to PAS and possesses 3′→5′ helicase activity, which can serve to unwind any duplex on the lagging-strand arm of the fork. Next, PriB joins PriA to form a PriA-PriB-PAS DNA complex. Then, DnaT-DnaB joins this complex and DnaB is transferred from the DnaT-DnaB complex to the DNA to form a *preprimosome* consisting of PriA, PriB, DnaT and DnaB on the DNA. Complete primosome assembly occurs when DnaG is added to this complex by virtue of a protein–protein interaction with DnaB. This primosome has been named the *replication restart primosome*.

In the case of *B. subtilis*, two checkpoints have been identified to the left and to the right of the *oriC* where the elongating DNA polymerases can be reversibly arrested. This arrest is caused as an effect of the stringent response (see Section 9.7) which is caused by amino acid starvation. These two checkpoints have been denoted as LSTer (*l*eft *st*ringent *ter*minus region) and RSTer and are located ~200 kb on either side of the *oriC*. Arrest at both of the checkpoints requires the replication

terminator protein RTP (see below), suggesting the existence of *Ter*-like sites at the checkpoints. Upon addition of amino acids to the cells, the forks are released from the checkpoints and continue duplication of the chromosome until arrested and terminated at the terminus region called *terC*.

3.1.1.4 Termination of Replication

Termination of replication of prokaryotic chromosomes occurs at specific sequences called *ter*mination sites (*ter*) exhibiting polarity, to which a protein binds which interacts with the unwinding helicase DnaB and impedes its action. Therefore, this protein has been termed *contrahelicase*. Termination of replication has been studied quite extensively in *E. coli* and *B. subtilis*.

Termination of Replication of the *E. coli* Chromosome In *E. coli*, there are ten *replication termini* (*ter*) located in a region diametrically opposite to the replication origin (Fig. 3.3A). The *ter* sites are 22 bp long and have polarity, i.e., they arrest replication forks, when they are present in one orientation with respect to *oriC*, but allow forks to pass through unimpeded in the opposite orientation. The *ter* sites are located in two clusters of five each, with each cluster having a polarity effect opposite to that of each other. Thus, the arrangement of the *Ter* sites forms a replication trap that forces the two forks, initiated at *oriC*, to meet each other within a well defined region of the chromosome. The *ter* sites specifically interact with the 36-kDa replication terminator protein called Tus (*t*erminus *u*tilization *s*ubstance; one monomer per *ter*), which impedes the DNA unwinding activity of DnaB in an orientation-dependent manner.

It is difficult to speculate as to why each chromosome has several termini, rather than the minimal requirement of two (one to stop the clockwise replication fork, the other to stop the counter-clockwise replication fork). It is possible that, because the *ter*-terminator protein complex is not capable of holding back the replication fork indefinitely, the multiple termini are needed to effectively coordinate fork arrest during a round of replication. However, the *Ter* sites are not essential, indicating that replication is terminated wherever the two forks happen to meet. This finding suggests that the Ter sites act as pause sites rather than as an absolute barrier to replication.

Termination of Replication of the *B. subtilis* Chromosome The *ter* sites of *B. subtilis* are bipatite and overlapping, with each inverted repeat consisting of both a ''core'' sequence and an ''auxiliary'' sequence (Fig. 3.3B). The core sequence binds to a dimer of RTP (for *r*eplication *t*ermination *p*rotein), a sequence-specific DNA-binding protein; and this protein–DNA complex cooperatively allows the binding of a second dimer of RTP to the auxiliary site. The auxiliary site does not bind to RTP in the absence of the core site and two interacting dimers of RTP are needed to effect replication fork arrest. The interaction of RTP (a homodimer, subunit molecular mass 29 kDa) with the *ter* sequence causes an about 30° bend in the DNA.

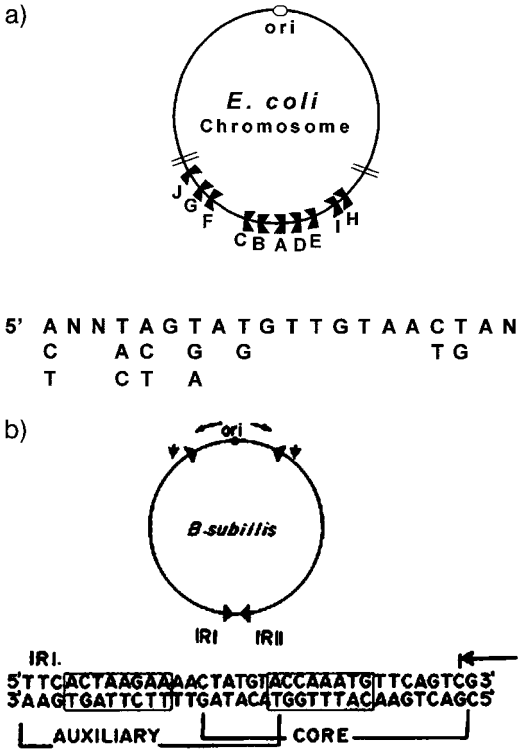


Fig. 3.3 (A) Termination of replication of the *E. coli* chromosome. The *E. coli* chromosome showing the *oriC*, the termination region with the location of the ten *ter* sites (upper part) and the consensus sequence of *ter* (lower part). S. Mulugu **2001**, *Proc. Natl. Acad. Sci. USA* 98, 9569; Fig. 1. (B) Termination of replication of the *B. subtilis* chromosome, showing the location of the replication checkpoints (ψ) and the two replication termination regions, IRI and IRII. The structure of the IRI sequence consists of two partially overlapping sequences, a core and and an auxiliary sequence. The two direct repeats present within IRI are boxed. D.E. Bussiere, D. Bastia **1999**, *Mol. Microbiol.* 31, 1611; Fig. 1.

3.1.1.5 **Separation of the Two Chromosomes**

Before the two daughter chromosomes segregate, they must be completely separated. But sometimes, the two chromosomes are joined together, either by forming a dimer or by being catenated. As the two converging replication forks approach each other at the termination region, removal of the final links by DNA gyrase action ahead of the forks becomes compromised by the restricted accessibility to enzyme in the short unreplicated region. Diffusion of these links behind the replication fork, to produce precatenanes, allows completion of replication, where the conversion of precatenanes to catenanes results in sister molecules being topologically entangled.

Mechanisms have been developed to ensure separation of the two chromosomes. Daughter chromosomes can become interlinked like the links of a chain for two reasons. Catenation can be:

1. the result of terminating a round of chromosome replication (after the two forks have merged, a two-step process involving the melting of the DNA helix and subsequent repair synthesis leads to an intermediate of two catenated daughter molecules);
2. caused by topoisomerases passing the strands of the two DNAs through each other.

Such interlinks are removed by breakage of both strands of one of the two DNAs and by passing the two strands of the other DNA molecule through the break. This enzymatic reaction is called *decatenation* and is catalyzed in *E. coli* by the topoisomerase IV, which efficiently removes precatenane and catenane links and may act preferentially in the replication termination region. By binding to the ParC subunit, SeqA stimulates the activities of topoisomerase IV (Fig. 3.4).

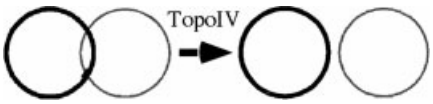


Fig. 3.4 Decatenation of two chromosomes. Topoisomerase IV recognizes catenated chromosomes and resolves them into two separated monomers. K.P. Lemon **2001**, *Genes Dev.* 15, 2031; Fig. 4A.

In addition to the presence of catenation links, the separation and segregation of newly replicated circular chromosomes can also be prevented by the formation of circular chromosome dimers, which can arise during crossing over by homologous recombination. Many circular genomes, both bacterial and plasmid, code for dimer-to-monomer resolution systems. Strains unable to resolve dimers display decreased growth rates and plating efficiencies resulting from frequent chromosome segregation defects, which results in the formation of filaments, subsequent induction of the SOS system and cell death. In *E. coli* where these dimers arise about once in every six generations as a consequence of recombination-dependent repair of collapsed replication forks, they are resolved to monomers by the action of the FtsK-XerCD-*dif* chromosome dimer resolution machinery. In short, two site-specific recombinases, XerCD, act at a 28-bp recombination site, *dif*, located in the terminus region to remove the crossover introduced by dimer formation, thereby converting dimers to monomers (see Section 4.2.2 for details of the reaction).

3.1.1.6 **Regulation of Chromosomal Replication**

All organisms have developed mechanisms that ensure chromosomal replication occurs once and only once per generation. In *E. coli*, three systems have been described that prevent reinitiation of origins for a given time window:

1. sequestration of *oriC* and of the *dnaA* promoter region;
2. binding of DnaA protein to a region close to *oriC* that provides a sink for DnaA;
3. regulatory inactivation of ATP-DnaA at the end of the initiation cycle.

Sequestration The *oriC* of *E. coli* contains an unusually high number of 11 GATC sequences (there are a total of 19 130 in the whole chromosome), within the 245-bp *oriC* (if statistically distributed, it should occur once per 256 bp), which are recognition sites for the Dam methylase. When the *oriC* region has been replicated, these GATC sites are present in a hemimethylated state for about 10 min when the doubling time of the cells is about 30 min, whereas elsewhere on the chromosome they become remethylated within approximately 1 min. The reason for the *oriC* region not being remethylated immediately is that these sequences are sequestered by cellular factors called SeqA and SeqB and form a *membrane-bound sequestration complex* which specifically interacts with hemimethylated DNA. SeqA localizes to so-called 'organizing centers' at midcell. How these 'SeqA foci' are built is not known, but these structures seem to be required for proper organization and/or segregation of daughter chromosomes. SeqA binds first to two sites in *oriC*, one on each side of DnaA box R1, from where it spreads by cooperative binding to adjacent regions. In addition, sequestration of the *dnaA* promoter region takes place, leading to a transient suppression of *dnaA* expression. SeqB does not bind to *oriC* independently, but enhances the ability of SeqA to do so in a hemimethylation-specific manner. A third protein, HobH (NapA) may also be involved in the formation of this membrane complex.

Titration of DnaA In the *E. coli* chromosome, there are 308 DnaA boxes that have variable affinity to DnaA. Among these DnaA boxes, five chromosomal regions show an especially high affinity with the *datA* locus exhibiting the highest one. The *datA* locus (for *D*na*A* *t*itration *A*) can bind about eight times more DnaA molecules than the *oriC* region with four DnaA boxes. The maximum binding capacity of the *datA* region was calculated to be about 300–400 DnaA molecules. This means that about one-third of the cellular DnaA content can be stored at this locus. That the *datA* locus is really involved in regulation of initiation has been shown by two different experiments. While deletion of *datA* results in overinitiation, additional copies of *datA* on a plasmid reduce initiation. In summary, all newly synthesized DnaA protein is bound to the DnaA boxes scattered around the chromosome. These boxes increase in number due to replication; and DnaA continues to be synthesized until the number of DnaA molecules exceeds the number of DnaA boxes, leading to initiation.

Inactivation of ATP-DnaA The DnaA protein is inactivated soon after initiation occurs. The DnaA protein can form a stable complex with ATP (K_D, 0.03 μM) or ADP (K_D, 1 μM); and only the ATP-bound form is active in initiation. Studies using replication cycle-synchronized cultures suggest that the number of ATP-DnaA molecules increases prior to initiation and the ATP-DnaA is converted to the inactive ADP form after initiation. This reaction is promoted *in vitro* by the β-subunit sliding clamp of the Pol III holoenzyme. As the β-subunit sliding clamp is formed upon loading of the replicase, DnaA becomes inactivated. This inactivation is directly linked to the initiation of replication and it coordinates switching from initiator activity to the replication process. The interaction between the β-

subunit and DnaA requires the DnaA-related protein Hda (*h*omologous to *DnaA*; 27 kDa). Hda belongs to the chaperone-like ATPase family, AAA^+, as does DnaA. Hda mediates the interaction of DnaA with the β-subunit sliding clamp to promote hydrolysis of DnaA-bound ATP. Interaction of Hda with β is mediated through the hexapeptide motif (QL[SD]LPL) present in the N-terminus of Hda. This rapid hydrolysis of ATP is proposed to be the main mechanism that blocks multiple initiations during cell cycle and acts as a molecular switch from initiation to replication.

The timing of initiation is tightly coordinated with the cell cycle progression. The ADP-DnaA protein may be reactivated in a cell cycle-dependent manner, which may determine the time-specific initiation. Acidic phospholipids such as cardiolipin have been suggested to reactivate ADP-DnaA by releasing bound ADP to be exchanged with ATP.

3.1.2
Replication of Linear Chromosomes and Plasmids

In general, linear chromosomes have a central *oriC* and their two ends are termed *telomers*. In *B. burgdorferi* and most *Streptomyces* species, replication proceeds bidirectionally from the *oriC*. Because of the polarity of replication, linear chromosomes are confronted with the problem of copying their 3′ ends. Bacterial linear chromosomes and plasmids have adopted at least two different strategies to replicate their 3′ ends based on the structure of the telomers. *Borrelia* chromosomes have covalently closed hairpin structures at their termini similar to those reported for *Borrelia* linear plasmids and the *E. coli* phage N15, which is maintained as a linear prophage with hairpin ends. Soil bacteria of the genus *Streptomyces* are unusual among bacteria in having protein-capped linear chromosomes. The terminal sequences of these chromosomes consist of *t*erminal *i*nverted *r*epeats (TIRs) of variable lengths, ranging from 20 kb to 550 kb. The nucleotide sequences of these TIRs are generally not conserved except for the first ~200 bp, which are packed with palindromic sequences. The covalently bound *t*erminal *p*roteins (TPs) are linked to the 5′ phosphate ends most probably by a phosphodiester linkage at a serine residue on the TP, based on the finding that the TPs can be easily removed by mild alkaline conditions. Similar telomeric structures have been reported in several bacteriophages, including *E. coli* phage PRD1 and *B. subtilis* phage Θ29.

Two models have been suggested to account for the replication of linear DNA with covalently closed hairpin ends, with near perfect inverted repeats at the two ends (Fig. 3.5). In pathway A, initiation of replication occurs internally, resulting in the generation of a circular dimer. Next, this double-stranded circular intermediate is processed by DNA breakage and reunion within the two telomeric regions, to generate covalently closed hairpin ends. In pathway B, initiation occurs by site-specific nicking in the telomeric repeats. Next, the 3′ hydroxyl ends at the nick site fold back and are used as primers in displacement DNA synthesis. Replication then generates two progeny molecules, which are subsequently sealed.

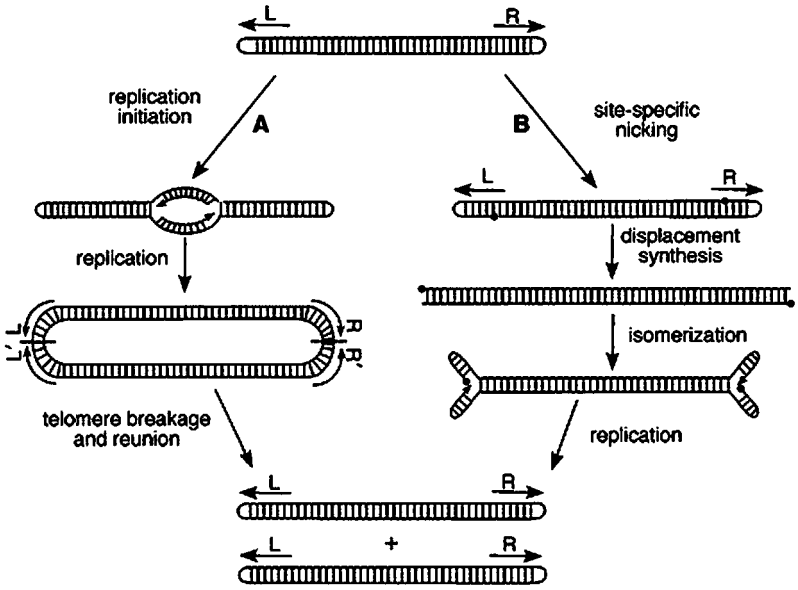


Fig. 3.5 Two models for replication of linear DNA with haipin ends. (A) Replication starts at a centrally located *oriC* and proceeds towards the ends leading to a circle dimer. An endonuclease attacks both telomers; and a DNA ligase circularizes the four ends. (B) Site-specific nicking generates two 3′ OH ends which are extended by Pol III, leading to tandem dimeric ends. These fold back to generate the new hairpin ends and serve as primers for replication of the complete molecules followed by ligation. G. Chaconas **2002**, *EMBO J.* 20, 3229; Fig. 1.

Both linear chromosomes and linear plasmids of *Streptomyces* are replicated bidirectionally from internal origins. This can leave single-stranded gaps at the 3′ ends of the replicons, which are patched by an as yet undefined mechanism, presumably involving the TPs. In *S. coelicolor*, the gene *tpgC* coding for the TP of 20 kDa is about 100 kb from the right end of the chromosome; and the close physical linkage between the *tpgC* gene and its cognate telomere favors their cosegregation and coevolution.

It should be mentioned that, at least in *Streptomyces*, linearity of the chromosome is a reversible state. Circularization occurs spontaneously at high frequency under laboratory conditions, generally deleting both telomers. Linearity in plasmids is also reversible and has been observed with plasmids from *Streptomyces* and *Borrelia*.

3.2
Partitioning (Segregation) of the Daughter Chromosomes

After duplication, the two daughter chromosomes have to be moved towards the poles of the elongated cell. This process is called segregation or partitioning and

seems to start long before the duplication of the genome reaches completion. The combination of experimental data with the analysis of the genome sequences derived from more than 130 bacterial species indicates that all bacteria use the same principles for chromosome organization and segregation, although details of how these processes are regulated and integrated into the lifestyle of a particular organism differ. Three different models have attempted to explain chromosome segregation in prokaryotic cells.

The Replicon Model

This model, proposed in 1963 by Jacob and coworkers, suggests that newly replicated sister chromosomes are attached to centrally located sites on the cell membrane that move towards opposite cell poles in parallel with cell elongation. Attachment of the chromosomes to the membrane is provided by the coupling between transcription and translation of genes specifying cotranslationally secreted proteins transiently anchoring DNA to the membrane, which in turn slowly pulls the chromosomes apart as the cell elongates. Chromosome segregation is essentially passive in this model.

The Replication Factory Model

This model suggests that spooling of DNA through a replication factory located at midcell might be an important driving force in chromosome segregation (Fig. 3.6A).

The Train-on-track Model

Here, independent sister replication forks move around the cell in a dynamic manner (Fig. 3.6B).

There are experimental data for both actual models underlining the dynamic behavior of the segregating chromosomes (see below).

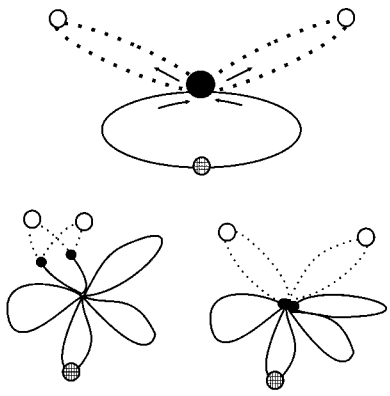


Fig. 3.6 Models for replication of the bacterial chromosome. (A) The factory model. This model predicts that the replication machinery remains in the middle of the cell, while the two origins move towards the cell poles. (B) The train-on-track model. In this model, the sister replication forks move freely within the cell.

Accurate chromosome partitioning is influenced by the concerted action of the following five mechanisms:

1. Replication may orient and push the origins in opposite directions.
2. *B. subtilis* and most probably also *E. coli* have a centrally located replication factory.
3. Condensation of the two chromosomes by supercoiling and condensins is important for accurate chromosome partitioning.
4. Homologs of plasmid partitioning proteins enhance chromosome partitioning.
5. Bipolar migration is mediated by *migS*.

Orientation of the Bacterial Chromosome Within the Cell

Soon after duplication, origin regions separate from each other and move to opposite sides of the cell. While those of *E. coli* and *C. crescentus* are found close to opposite cell poles, the *oriC* of *B. subtilis* is located near a position that will become midcell following cell division. The two new copies of the terminus region stay in midcell until medial cell division and then move towards the midcell in the two daughter cells. How this movement occurs is unknown and may involve specific terminus information. Also unknown are the molecular details of how the origin and terminus regions are positioned. Since both regions are enriched in membrane fractions, both could be membrane-anchored.

The Replication Factory

E. coli has a centrally located replication factory. Newly duplicated DNA is located at the cell center in the majority of the cells, through which the DNA slides and is pushed out towards the cell poles. This finding is corroborated by the localization of SeqA which forms either one centrally located focus or two foci, one in each half of the cell. Although SeqA also binds specifically to the *oriC* region to regulate initiation of replication (see above), these foci do not colocalize with the *oriC*. In these foci, the bulk of the ~1000 SeqA molecules bind at GATC sites to newly duplicated hemimethylated DNA that has just exited the replisome during the few minutes before methylation of these sequences occurs through the action of the Dam methylase. Two very different models have been proposed for how the two replicated DNA molecules segregate into daugther cells: the *factory model* and the *train-on-track model* (Fig. 3.6). In the first model, the replisome pulls the DNA template into the cell center, duplicates it and then releases the product into the opposite halves of the cell. It has been suggested that the movement away from the replication factory is assisted by proteins such as SeqA in *E. coli* and Spo0J-*parS* in *B. subtilis* (see below). The terminus segregates after termination of replication and resolution of the topologically linked sister chromosomes. In the second model, the newly replicated chromosomes move around within the cell in a dynamic manner.

Is There a Bacterial Centromere?

How do the origin regions migrate to the cell poles during the bacterial cell cycle? Since their positioning does not depend on their presence *per se*, another *cis*-acting region has to be involved. Such a region has been identified as being responsible for the bipolar positioning of *oriC* in *E. coli* and is termed *migS*, a 25-bp sequence comprising an imperfect stem-loop. Studies are underway to identify the protein(s) interacting with *migS*. A functionally related site is present in *B. subtilis* during sporulation. Here, the RacA protein binds preferentially near the *oriC* and interacts with the DivVA protein, a membrane protein, localized preferentially at the poles. A putative RacA-binding motif has been identified consisting of a 14-bp hairpin.

Condensins and Chromosome Compaction

Chromosome compaction and organization are important for partitioning. This condensation is based on two principles: supercoiling (see Section 2.1.1) and the activity of proteins called *condensins*. Two different classes of condensins have been identified, called SMC (for *s*tructural *m*aintenance of *c*hromosomes) and MukB, which are functional analogs. SMC homologs have been found in nearly all Archaea and Gram-positive bacteria and in about 40% of the Gram-negative bacteria with partially sequenced genomes. More than half of the Gram-negative bacterial species lack a *smc* gene. Instead, a variety of γ-proteobacteria contain the MukB protein. Electron micrographs of *B. subtilis* SMC and and *E. coli* MukB revealed similar structures (Fig. 3.7): two globular domains separated by a long rod-like domain with a flexible hinge in the middle. Condensins have been shown to

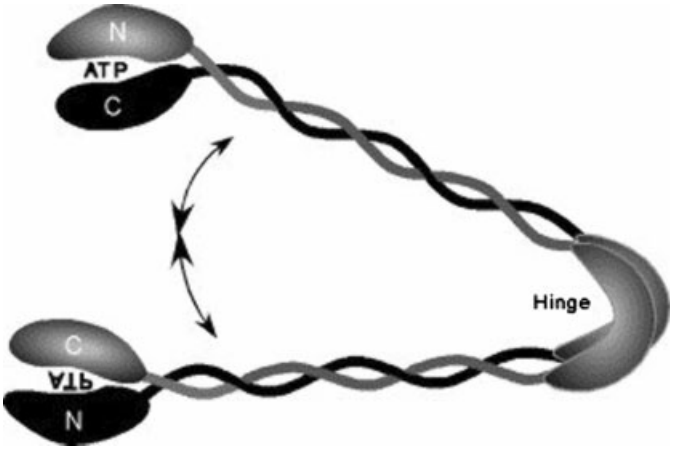


Fig. 3.7 Schematic structure of SMC proteins. Two monomers form an interwound dimer in an antiparallel fashion, leading to an interaction of the N- and C-termini at both ends which binds ATP. A hinge region in the middle of the dimer allows flexibility. P.L. Graumann **2001**, *Biochimie* 83, 53–59; Fig. 1.

bind DNA; and it has been suggested that the function of condensins is to organize the chromosome into a highly ordered structure by constraining supercoils. Electron micrographs of *B. subtilis* SMC showed that it dimerizes in an antiparallel fashion. Antiparallel dimerization results in totally symmetrical molecules and brings together the N- and C-terminal globular domains, which contain the highly conserved nucleotide-binding motifs Walker A and Walker B, respectively, to form the SMC head domain with the capacity to bind DNA and hydrolyze ATP. Self-folded SMC monomers associate via hinge–hinge interactions to form V-shaped dimers that represent the functional SMC unit.

B. subtilis smc mutants are not viable in rich media at 37°C but can grow at 23°C. They display several typical phenotypes, even at the permissive temperature, such as: (a) very frequent formation of anucleate cells, (b) splitting of the nucleoid by septum in dividing cells (guillotine effect), (c) frequent cell elongation and (d) extension and anomalous posititioning of the nucleoid in elongated cells. The partitioning defect observed in *smc* and *mukB* mutants is probably a secondary consequence caused by defects in chromosome organization and compaction.

Condensins interact with additional proteins, where the MukB protein forms a complex with two proteins essential for chromosome partitioning, MukE and MukF, and all three genes form a tricistronic operon. Proteins interacting with SMC have been designated as Scps (for *s*egregation and *c*ondensation *p*roteins). The genes *scpA* and *scpB* are localized next to *smc* genes in nearly all SMC-containing archaea, suggesting that they belong to the same operon and are thus involved in a common process in the cell. The ScpA N- and C-terminal domains are conserved in non-SMC subunits of eukaryotic cohesin complexes, thus defining a new family of *kleisin* proteins. *scpA* mutants display characteristic phenotypes nearly identical to those of *smc* mutants. Thus, both SMC and ScpA are required for chromosome segregation and condensation. Interestingly, mutants of the *B. subtilis* gene *scpB*, localized downstream from *scpA*, display the same phenotypes, which indicate that ScpB is also involved in these functions.

Par Proteins and Origin Region Partitioning

Partitioning of low-copy plasmids is accomplished by Par systems. The Par system was first identified with P1 (*parABS*) and F (*sopABC*). The two *par* genes are in an operon adjacent to the *cis*-acting site, *parS*. While ParA is a Walker-type ATPase and a DNA-binding protein functioning as a transcriptional repressor of the *parAB* operon, ParB binds to *parS* and modulates both ATPase and transcriptional repressor activity of ParA. Both F and P1 plasmids localize to the midcell and after replication, to both cell quarter positions. Since ParA is membrane-associated, it might tether the ParB-*parS* complex at specific cellular positions to a putative host cell receptor.

Chromosomal homologs of the plasmidal Par system have been identified in many bacteria. In *B. subtilis*, Spo0J (ParB) binds to at least eight *parS* sites located in the origin proximal 20% of the chromosome, forming a large nucleoprotein complex. This complex can be made visible by either immunofluorescence or a

functional Spo0J-GFP fusion. One single Spo0J focus is present per origin region, suggesting that Spo0J organizes that region by bringing most of these *parS* sites together. Spo0J is somehow involved in accurate partitioning of the *B. subtilis* chromosomes, since *spo0J* null mutants produce 1–2% anucleate cells. In addition, overreplication occurs in the absence of Spo0J, suggesting that this protein might influence the frequency of initiation of DNA replication. There is also a *parA* homolog termed *soj* (for suppressor of *spo0J* during sporulation), but in its absence, there is no chromosome partitioning defect, while the Soj protein has a clear role as a negative regulator of transcription during sporulation. In *C. crescentus*, ParA and ParB are present. During most of the cell cycle, both proteins are equally distributed throughout the cytoplasm. Immediately before cell division, they concentrate at the cell poles, where they either hold the chromosomes during cell division or help to pull them apart.

Postseptational Partitioning

Many bacterial species seem to encode for a mechanism allowing postseptational partitioning. This mechanism ensures that any chromosome that might be caught in the newly formed septum can still be successfully partitioned. The FtsK protein of *E. coli* and the SpoIIIE protein of *B. subtilis* are members of a family of proteins localized to the leading edge of the septum and function as ATP-dependent DNA pumps. It has been suggested that these proteins, in an ATP-dependent manner, move chromosome(s) out of the closing division septum.

Important open questions:

- What are the molecular details involved in positioning the origin and terminus regions?
- What information is responsible for the movement of the two terminus regions from the cell poles to a midcell position after medial cell division?

3.3 Cell Division

The process of cell division is one of the most complex and least understood phenomena in the cell biology of bacteria. Almost all cells divide at a highly reproducible location within the cell. Therefore, cells have first to find their appropriate division site and then localize their division machinery to that site. This process is best characterized in *E. coli*, where the three proteins MinC, MinD and MinE collaborate to position the cell division protein FtsZ at the midcell. Next, nine essential proteins localize to the septum and form the *divisome*. These proteins coordinate invagination of the cellular membrane, inward growth of the peptidoglycan layer and, finally, promote separation of the two daugther cells. All these proteins and their known functions are summarized in Table 3.3.

The predicted topology of the ten known *E. coli* cell division proteins is shown in Fig. 3.8. What is known about the proteins involved in cell division?

Table 3.3 Genes involved in *E. coli* cell division.

Gene	*Protein size (kDa)*	*Function*
ftsA	45.3	Cytoplasmic protein of the AAA^+ family
ftsB		Bitopic membrane topology; might form a heterodimer with FtsL
ftsK	146.6	Polytopic membrane topology
ftsI	63.9	Bitopic membrane topology; penicillin-binding protein (PBP3) involved in peptidoglycan synthesis most probably at the septum
ftsL	13.6	Bitopic membrane topology; might form a heterodimer with FtsB
ftsN	35.0	Bitopic membrane topology
ftsQ	31.4	Bitopic membrane topology; may be involved in cell wall synthesis
ftsW	46.0	Polytopic membrane topology; homolog of RodA
ftsZ	40.3	Homologous to tubulin; forms a ring-like structure (Z-ring) by GTP-dependent polymerization at the site of cell constriction; interacts through its C-terminus with FtsA and ZipA
minC	25.4	Inhibits Z-ring formation at the cell poles; forms a complex with MinD
minD	29.6	Activates MinC; forms a complex with MinC
minE	9.7	Forms a ring-like structure and shields the midcell site from MinCD
zipA	36.5	Integral membrane protein with a C-terminal FtsZ-binding domain

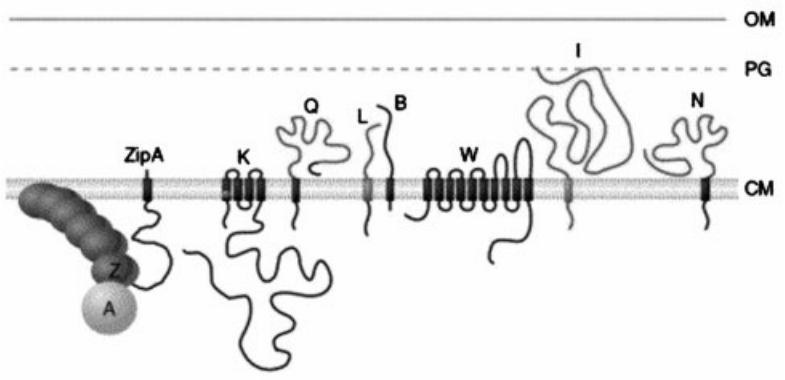


Fig. 3.8 Schematic representation of the topology of the ten proteins involved in cell division. While most Fts proteins are bitopic proteins, FtsK contains four and FtsW even ten transmembrane domains. N. Buddelmeijer, J. Beckwith **2002**, *Curr. Opin. Microbiol.* 5, 553–557; Fig. 1.

FtsA (*f*ilamentation *t*emperature-sensitive *A*) is a cytoplasmic actin-like phosphoprotein which belongs to the AAA^+ superfamily and is peripherally associated with the membrane. FtsA is typical of the actin superfamily, including sugar kinases and Hsp70 proteins, possessing two domains with a common core and an interdomain cleft harboring the nucleotide-binding site. FtsA forms dimers and

assembles at the Z-ring early; and the extreme C terminus of FtsZ seems to be the site of interaction with FtsA. The cellular ratio of FtsA to FtsZ is 1:100, with the implication that there is not sufficient FtsA to form a complete ring. It has been suggested that FtsA makes only widely interspersed contacts with FtsZ filaments. One specific role for FtsA is the recruitment of other cell division proteins to the division site. FtsA could use ATP hydrolysis to drive assembly or, alternatively, ATP hydrolysis could trigger a conformational change after binding to the Z-ring, thereby stabilizing this interaction.

FtsB contains a leucine zipper-like motif in the periplasmic domain, with the potential to form a coiled-coil structure. Since it colocalizes with FtsL to the septum, both proteins may form heterodimers or higher order oligomers and function as a unit in cell division.

FtsK, the largest protein of this group, is a 1329-amino-acid protein that is essential for septation and appears to have three domains. The 202-amino-acid N-terminal hydrophobic domain (domain I) localizes the protein to the septum in an FtsZ-dependent manner and is essential for cell division. The ~480-amino-acid C-terminal cytoplasmic domain (domain III) is implicated in chromosome segregation (resolution of chromosome dimers during DNA segregation; see Section 4.2.2) and contains a presumptive ATP-binding site. The C- and N-terminal domains are separated by a region of ~700 amino acids (domain II) which is proline- and glutamine-rich. FtsK exhibits homology with the *B. subtilis* SpoIIIE protein, which is implicated in the transport of chromosomal DNA into the prespore, and proteins implicated in plasmid conjugal transfer in Gram-positive bacteria. The N-terminal domain is predicted to contain four membrane-spanning regions which are likely to localize FtsK to the division site. Localization of FtsK to the septum is dependent on the prior localization of FtsZ and FtsA.

FtsI, also known as penicillin-binding protein 3 (PBP3; ~100 molecules per cell), is a bitopic membrane protein. The N-terminal transmembrane span (about 50 amino acid residues) is followed by a domain of about 200 amino acids of poorly understood function and then by a C-terminal domain of about 300 amino acids, which contains the characteristic sequence motif representing the catalytic residues of PBPs, the penicillin-binding domain. FtsI may act in a multiprotein complex that introduces three new glycan strands in parallel with hydrolysis of an existing docking or template strand. Genetic and biochemical evidence indicate that FtsI is required specifically for synthesis of peptidoglycan at the division septum, while PBP2, a homolog of FtsI, appears to be the primary transpeptidase for cell elongation.

FtsL, *FtsN* and *FtsQ* are three additional bitopic, membrane-anchored proteins localized at the midcell in constricting cells, which are essential for cell division but have otherwise completely unknown functions. FtsL is a small transmembrane protein with a poorly conserved primary sequence. The structure of FtsL includes a small N-terminal cytoplasmic domain (37 residues), a single transmembrane segment (20 residues) and a small periplasmic domain (64 residues). The sequence of the periplasmic domain of FtsL exhibits a stretch of five leucines separated from each other by seven residues, which has features of an α-helical leu-

cine zipper. The presence of this motif raises the possibility that FtsL interacts with itself or other proteins to form homo- or heterodimers. FtsQ (~22 copies per cell) consists of a short (24-amino-acid) cytoplasmic N-terminus, a single (25-amino-acid) transmembrane segment, and a relatively large (227-amino-acid) periplasmic domain. The FtsN protein has an N-terminal transmembrane segment that directs the major C-terminal domain to the periplasmic side of the membrane and exhibits weak sequence similarity to cell wall amidases, suggesting a possible role in wall hydrolysis. It could be required to cleave certain bonds in the wall to allow initiation of constriction. FtsQ could be detected only at the midcell in cells with a visible constriction. It may be involved in cell wall synthesis, because it is found only in bacterial species having a cell wall. Therefore, this protein seems to be required at a later stage in the cell division process.

FtsW is a polytopic integral membrane protein of the SEDS family of proteins (for *s*hape, *e*longation, *d*ivision, *s*porulation). Proteins of this family appear to have ten transmembrane segments and the most likely function for SEDS proteins seems to be translocation of the lipid-linked precursor for peptidoglycan synthesis from the inside to the outside of the membrane and delivery to its cognate peptidoglycan complex.

FtsZ, the best characterized of these proteins, is highly conserved in prokaryotes from *E. coli* to Mycoplasmas, which do not possess peptidoglycan in their envelope, to Archaea (*Haloferax volcanii*) and was also recently discovered in the cyanobacteria *Syneccocystis* sp. PCC6803. In addition, sequence homology has been found between FtsZ and a protein of *Arabidopsis*, which could have a role in chloroplast division. FtsZ is an ancestral homolog of the eukaryotic tubulin. The amino acid sequence identities of tubulin and FtsZ are less than 15%, much less than the 25% that is generally used as the gold standard for establishing homologs, but it exhibits limited, but significant sequence similarity to eukaryotic tubulins around the glycine-rich motif (GGGTGTG). Like tubulin, FtsZ has GTPase and polymerization activities and may provide the force necessary to constrict the cell. FtsA and ZipA both interact directly with FtsZ and ZipA stabilizes polymerized FtsZ. FtsZ polymerizes into the *Z-ring*, a cytoskeletal element formed through self-assembly, which localizes septal growth. Formation of the Z-ring is inhibited by the SulA (also called SfiA; see Section 5.2.9). The ring is present throughout septation, its diameter decreasing, and FtsZ is found dispersed in the cytoplasm before and at the completion of division. Recently, the crystal structure of FtsZ from *M. janaschii* was determined: FtsZ has two domains arranged around a central helix and GTPase activity is localized in the N-terminal part of the protein, whereas the C-terminal domain function is still unknown.

MinC, *MinD* and *MinE* cooperate to ensure the midcell positioning of the Z-ring by suppressing septation at aberrant sites located near the cell poles. They ensure that the Z-ring is assembled at the midcell and not at polar sites by masking the polar sites, leaving the midcell site accessible to FtsZ. The three genes encode a bipartite inhibitor MinCD that is topologically regulated by the third product, MinE. MinE consists of at least two domains, an N-terminal domain that counteracts the inhibitory activity of MinCD and a C-terminal domain that bestows topo-

logical specificity to MinE action. In wild-type cells, all three proteins assemble into a membrane-associated polar zone that grows toward the midcell. When the polar zone approaches the midcell, a ring of MinE appears at its medial edge. The polar zone then shrinks back to the cell pole, accompanied (pushed?) by the MinE ring. At the same time, the Min proteins assemble into a new polar zone at the opposite cellular pole, etc.

MinE exerts two functions:

1. It is required for formation of the MinCDE polar zone.
2. It forms the MinE ring oscillating from one pole to the other.

This oscillatory cycle is repeated many times within each cell cycle and keeps MinC division inhibitor away from the midcell to inhibit cell division at the polar division sites. Analysis of MinCD has revealed that MinC is an inhibitor of division that is activated by MinD. MinE might inactivate MinCD by inducing dissociation of this inhibitor.

ZipA (*Z i*nteracting *p*rotein *A*), a membrane-anchored protein, has been discovered from a search for proteins that interact directly with FtsZ. It contains three domains: a short N-terminal membrane-anchored domain, a central P/Q domain that is rich in proline and glutamine and a C-terminal domain, which comprises almost half of the protein (residues 185–328) and appears to be involved in interaction with FtsZ. Requirement of ZipA for cell division in *E. coli* was demonstrated by the formation of nonseptate filaments after overexpression or depletion of ZipA, indicating a ZipA-induced block in septum formation and suggesting that disruption of the ZipA–FtsZ interaction disrupts cell division.

Divisome Assembly

The first step in cell divison involves the FtsZ protein, which polymerizes into the Z-ring at the cell center between the two daughter chromosomes, which is also the future division site. If assembly of the Z-ring at midcell is prevented, anomalous divisions occur near the cell poles, indicating that the two cell poles are capable of supporting cytokinesis. The Z-ring is thought to serve as a framework for the assembly of the cell division machinery. In the absence of FtsZ, none of the other proteins are seen at the midcell. Formation of Z-rings at the cell poles is prevented by the polar presence of MinC, an inhibitor of FtsZ polymerization. MinC localization is dynamic and completely dependent on MinD. MinC and MinD rapidly oscillate from one pole to the other in the timespan of ~20 s, resulting in the lowest concentration of MinC at midcell. Oscillation of both proteins depends on MinE. The bulk of MinCD is assembled into spirals at one pole, while faint spirals of MinCD can also be detected at the opposite pole (Fig. 3.9). MinE appears to consist of one or two loops of a coiled structure; and faint MinE spirals can be detected extending from the midcell to the cell poles. It has been suggested that all three Min proteins are part of the same higher order structure and that their oscillations reflect a redistribution of subunits across a permanent scaffold of Min proteins.

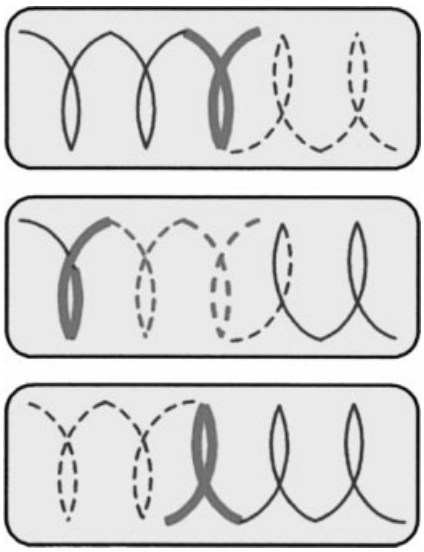


Fig. 3.9 Organization of the MinC, MinD and MinE protein within the *E. coli* cell. MinC and MinD (faint lines) and MinE (bold line) are organized into spiral structures. Movement of MinE from one pole to the other pushes the MinCD complex which oscillates from pole to pole. Z. Gitai **2003**, *Proc. Natl. Acad. Sci. USA* 100, 7423–7424; Fig. 1.

Invagination of the septum (constriction) may be driven by the force generated by depolymerization of the Z-ring following GTP hydrolysis. Formation of the Z-ring initiates at a single nucleation site and expands bidirectionally around the circumference of the cell. Through its C-terminus, FtsZ interacts with two other cell division proteins, FtsA and ZipA, which localize to the septum independently of each other. FtsK localization is dependent on FtsA and ZipA followed by FtsQ, whose appearance at midcell requires the presence of FtsK. FtsB and FtsL follow FtsQ in this sequential pathway and require each other for localization. FtsB stabilizes FtsL, probably through a direct interaction. FtsW localization is dependent on both FtsL and FtsB, followed by FtsI. At the end, FtsN completes the divisome. No observable initiation of cell division takes place until all the cell division proteins have been assembled. This observation indicates that, in the absence of just one of these proteins, aberrant cell division is prevented.

By what mechanism do most cell division proteins localize to the midcell in a sequential manner? Two models or a combination of them have been proposed. According to the first model, each protein, after localization to the division side via protein–protein interactions, undergoes an alteration in its structure to provide a template for the next protein to interact with. The second model claims that each protein, after recruitment to midcell, modifies that region, e.g., by changing membrane lipid composition or peptidogylcan structure. This alteration provides a structure recognized by the next protein in the pathway.

Open questions:

- Are there additional genes still to be discovered important for the cell division process?
- What are the exact functions of all the proteins assembled at the divisome?
- What is the sequence of reactions leading to formation of the septum and finally cell division?

3.4
Plasmid- and Chromosome-encoded Toxin–Antitoxin Modules

Plasmids and chromosomes can encode modules consisting of a pair of genes encoding a stable toxin and an unstable antitoxin that are coexpressed (so-called *t*oxin–*a*ntitoxin (TA) or ''addiction'' modules). Their expression is autoregulated either by a complex formed by toxin and antitoxin or by antitoxin alone. When coexpression is inhibited, the antitoxin is rapidly degraded by an ATP-dependent protease, enabling the toxin to act on its intracellular target. TA cassettes have a characteristic organization in which the gene for the antitoxin component precedes the toxin gene (Fig. 3.10). While the toxin gene codes for a stable protein,

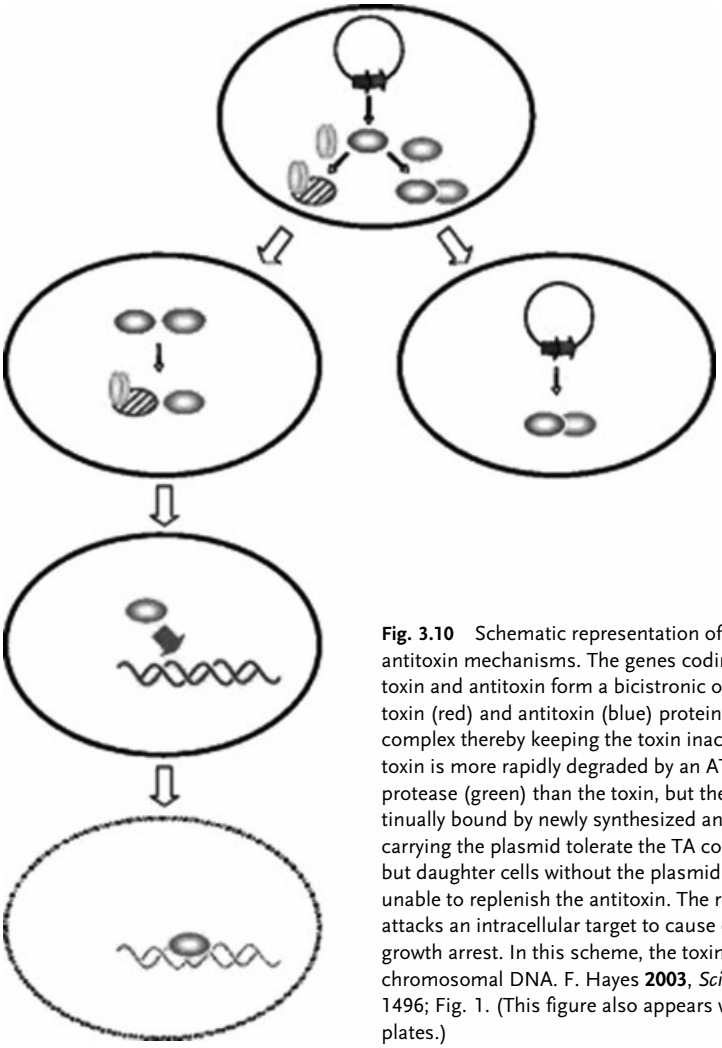


Fig. 3.10 Schematic representation of the toxin–antitoxin mechanisms. The genes coding for the toxin and antitoxin form a bicistronic operon. The toxin (red) and antitoxin (blue) proteins form a tight complex thereby keeping the toxin inactive. The antitoxin is more rapidly degraded by an ATP-dependent protease (green) than the toxin, but the latter is continually bound by newly synthesized antitoxin. Cells carrying the plasmid tolerate the TA complex (right), but daughter cells without the plasmid (left) are unable to replenish the antitoxin. The released toxin attacks an intracellular target to cause cell death or growth arrest. In this scheme, the toxin attacks the chromosomal DNA. F. Hayes **2003**, *Science* 301, 1496; Fig. 1. (This figure also appears with the color plates.)

the antitoxin is either a labile protein (type II module) or an untranslated antisense RNA species (type I module). In the latter case, the antisense antitoxin RNA inhibits translation of the toxin gene. If the antitoxin is a protein, it neutralizes the toxin by protein–protein interaction. The TA systems were originally discovered on low-copy-number plasmids where their function is to stabilize the plasmid carrying them by selectively killing daughter bacteria that have lost the plasmid. Database searches revealed that plasmid-encoded TA systems have homologs in the chromosomes of Eubacteria and Archaea. For instance, the *E. coli* chromosome encodes at least five TA systems: *relBE*, *mazEF* (*chpA*), *chpB*, *dinJ-yafQ* and *yefM-yoeB*.

3.4.1 Postsegregational Killing of Plasmid-free Cells

The segregational stability of low-copy-number plasmids is achieved by the activity of plasmid-specified partitioning proteins that direct plasmid copies to new daughter cells at cell division (see above). In addition, these plasmids code for mechanisms which result in selective killing or growth impairment of cells that have failed to acquire a plasmid copy. These mechanisms confer an advantage on plasmid-retaining cells by reducing the competitiveness of their plasmid-free counterparts and thereby ensure the retention of the plasmid in the population. These toxin–antitoxin mechanisms, also called *postsegregational killing systems*, attack cells from within.

The best studied type I module is the *hok-sok* locus of plasmid R1 of *E. coli*. The *hok* toxin gene codes for a translationally inactive transcript that is activated by by 3′-end processing. But translation of the processed transcript is inhibited by binding of the *sok* antisense RNA that is complementary to the 5′ end of the *hok* transcript. The double-stranded RNA is then degraded by RNase III. The half-life of the *hok* transcript is longer than that of the *sok* one. Therefore, in plasmid-free cells, processed *hok* mRNA accumulates which is translated into the 52-amino-acid Hok toxin. The Hok protein causes cell membrane depolarization. The majority of plasmid-encoded TA loci belong to type II; and the antitoxin is more susceptible to degradation by ATP-dependent proteases such as Lon and Clp (see Section 7.5.1). The target of the CcdB toxin encoded by the F factor is the catalytic GyrA subunit of gyrase and entraps a cleavage complex between gyrase and the DNA molecule. DNA and RNA polymerases are unable to traverse the trapped CcdB–gyrase complex. In addition, CcdB also binds to free gyrase to produce a complex impaired in supercoiling. Both interactions of CcdB with DNA gyrase are lethal for *E. coli*. The ParE toxin of plasmid RK2, though unrelated evolutionarily to the CcdB toxin, poisons the activity of gyrase in a manner strikingly similar to that of CcdB. The last example to be mentioned here is Kid toxin encoded by plasmid R1. This toxin specifically inhibits DnaB-dependent replication, either by blocking assembly of the replication complex or by interfering with DnaB-mediated protein interactions necessary for the formation of a productive complex.

Postsegregational killing by TA modules is a highly effective strategy that plasmids use to ensure their persistence within the bacterial population.

3.4.2 Chromosomal Toxin–Antitoxins

Bacterial chromosomes also harbor TA modules that are homologous to those identified on plasmids but fulfill a different function. The *E. coli* K12 chromosome contains several type I TA cassettes related to *hok-sok* but apparently inactive. In addition, the chromosome includes a number of type II TA systems that either enhance segregational stability and/or exhibit toxic–antitoxic behavior when inserted ectopically into multicopy plasmids such as the *mazEF* (also called *chpA*), *chpBIK* and *relBE* loci. The *mazEF* locus was described first, where *mazF* codes for the toxin and *mazE* for the unstable antitoxin which is degraded by the ClpAP serine protease. Both proteins form a linear heterohexamer composed by alternating toxin and antidote homodimers ($MazF_2$-$MazE_2$-$MazF_2$). Native MazF protein cleaves RNA at the 5′ side of residues A in 5′-NAC-3′ sequences (where N is preferentially U or A). MazF-dependent cleavage occurs at target sequences situated either in single- or double-stranded RNA regions. It has been suggested that the substrate specificity of the MazF endoribonuclease may be modulated by additional factors to cleave messenger and other cellular RNAs. Blockage of *mazEF* transcription and translation inhibits *de novo* synthesis of the unstable antitoxin; and the uncomplexed MazF attacks its so far unknown intracellular target, leading to an impairment of translation and/or DNA replication. Expression of the *mazEF* locus is triggered by the amino acid starvation signal ppGpp (see Section 9.7), by thymine deprivation and by antibiotics that act as general inhibitors of transcription or translation. The detrimental effect of the MazF toxin can be fully reversed when its cognate antitoxin is produced. The *relBE* locus is equally activated in response to amino acid starvation and elevated levels of ppGpp. The RelE toxin is an inhibitor of translation which cleaves transcripts within the ribosome (but not free transcripts) and induces a bacteriostatic response that can be reversed by synthesis of the RelB antitoxin. After cleavage, the ribosomes are stalled and subsequently released by binding of the tmRNA (see Section 6.4.2.2). It has been suggested that both TA cassettes are modulators of the physiologial response to poor nutritional conditions.

Suggested Reading

D.E. Bussiere, D. Bastia **1999**, Termination of DNA replication of bacterial and plasmid chromosomes, *Mol. Microbiol.* 31, 1611–1618.

J. Errington, et al. **2003**, Cytokinesis in bacteria, *Microbiol. Mol. Biol. Rev.* 67, 52–65.

F. Hayes **2003**, Toxins–antitoxins: plasmid maintenance, programmed cell death and cell cycle arrest, *Science* 301, 1496–1499.

K. Kobryn, G. Chaconas **2001**, The circle is broken: telomere resolution in linear replicons, *Curr. Opin. Microbiol.* 4, 558–564.

K.P. Lemon, A.D. Grossman **2001**, The extrusion–capture model for chromosome partitioning in bacteria, *Genes Dev.* 15, 2031–2041.

W. Margolin **2000**, Themes and variations in prokayotic cell division, *FEMS Microbiol. Rev.* 24, 531–548.

D. RayChaudhuri, et al. **2001**, Protein acrobatics and bacterial cell polarity, *Proc. Natl. Acad. Sci. USA* 98, 1332–1334.

L. Rothfield, et al. **1999**, Bacterial cell division, *Annu. Rev. Genet.* 33, 423–448.

J. Sawitzke, S. Austin **2001**, An analysis of the factory model for chromosome replication and segregation in bacteria, *Mol. Microbiol.* 40, 786–794.

4
Recombination

Genetic recombination is a fundamental cellular process in biology that continuously rearranges genes or part of genes both within and between the chromosomes of all organisms and is defined as the breakage and rejoining of DNA into new combinations. Through recombination events, the genes within and between chromosomes may be rearranged, segregation at cell division may be modulated and DNA repair facilitated. This process may involve the same DNA molecule (*intramolecular* recombination) or two different molecules (*intermolecular* recombination). Three different mechanisms of recombination have been described: *homologous*, *site-specific* (also called *sequence-specific*) and *illegitimate* recombination.

4.1
Homologous Recombination

Homologous recombination is a fundamental cellular process ubiquitous in all organisms and involves exchanges between DNA molecules (or parts thereof) of identical or nearly identical sequence for considerable distances along their length. It provides therefore a potent evolutionary force that serves both to promote genetic diversity and to conserve genetic identity. Recombination between two regions of two DNA molecules can also occur between similar DNA sequences that occur in more than one place along the DNA. This type of recombination is called *ectopic* or *homeologous* recombination and can lead to integration of plasmid DNA, deletions and inversions. In particular, homologous recombination is involved in the following processes:

- rearrangement of genes within and between chromosomes
- promotion of DNA repair
- segregation of chromosomes at division
- restart of blocked or incomplete replication forks.

Homologous recombination can occur by a number of different mechanisms, all of which lead to the formation of heteroduplex DNA. In *E. coli*, two major pathways (recombination machines) of homologous recombination operate: the *RecBCD* and the *RecF* pathways. The RecBCD pathway is the main route when a

Dynamics of the Bacterial Chromosome. Structure and Function. W. Schumann
Copyright © 2006 Wiley-VCH Verlag GmbH & Co. KGaA
ISBN 3-527-30496-7

DNA nick is encountered and the RecF pathway seems to act when a DNA base lesion is present. Both recombination machines consist of at least three broad classes of activities: helicases, nucleases and synapsis proteins. The RecBCD and RecF recombination machines use distinct sets of enzymes to produce a single-stranded DNA molecule coated with RecA protein. In the RecBCD pathway, all the required functions reside in one machine: the helicase, nuclease and RecA-loading activities. In contrast, the RecF pathway requires several separate proteins (see below).

The *recA* gene was the first to be identified as being involved in homologous recombination: it regulates multiple repair pathways via its role in the induction of the SOS response (see Section 5.2.9). *recA* null mutants exhibit a reduction in their recombination frequency by 10^{-3}. Null mutants in *recB* or *recC* are deficient in recombination and these mutant cells are also hypersensitive to UV and X-rays

Table 4.1 *E. coli* enzymes involved in homologous recombination.

Enzyme	*Size of protein (kDa)*	*Function*
Presynapsis		
RecBCD	134; 129; 67	ATPase; 5′→3′ dsDNA exonuclease; DNA helicase; χ-specific endonuclease
RecE	96	5′→3′ dsDNA exonuclease
RecJ	63	5′→3′ ssDNA exonuclease
RecQ	68	ATP-dependent 3′→5′ helicase
HelD	22	helicase IV
Synapsis		
RecA	38	Forms helical filaments on ssDNA; ATPase; catalyzes homologous pairing and strand exchange
RecF	40.5	ssDNA binding protein
RecO	27	Stimulates binding of RecA to ssDNA (with RecR)
RecR	22	Stimulates binding of RecA to ssDNA (with RecO)
RecT	30	ssDNA binding; promotes DNA pairing
SSB	18	ssDNA binding protein; melts secondary structures in ssDNA
Postsynapsis		
RuvA	22	Binds specifically to Holliday junctions; targets RuvB to DNA
RuvB	37	ATPase; catalyzes branch migration with RuvA
RuvC	19	Endonuclease; cleaves specifically Holliday junctions
RecG	76	ATPase; binds to Holliday junctions and promotes branch migration
RusA	14	Endonuclease; cleaves specifically Holliday junctions
LigA	54	Seals DNA nicks

and have reduced plating efficiencies. In contrast, *recD* null mutants are deficient in DNA degradation while their recombination, DNA repair and cell viability are similar to wild-type bacteria. The RecF pathway was identified by screening mutagenized *recBC* cells to isolate the genes required for the low recombination remaining in these cells and led to the detection of *recF*. However, when the *recF* mutations were moved into a *recBC*$^+$ background, the recipient cells were found to have normal levels of recombination, although they were still sensitive to UV and displayed a low plating efficiency. These and some other observations link the RecF pathway to DNA replication. The proteins known to be involved in homologous recombination are summarized in Table 4.1.

4.1.1 The RecBCD Pathway of Homologous Recombination

Homologous recombination involving the RecBCD pathway is divided into the three steps *presynapsis*, *synapsis* and *postsynapsis*. The presynapsis step provides single-stranded DNA catalyzed by the RecBCD complex that is subsequently recognized and bound by the RecA protein forming a nucleoprotein complex (synapsis). This complex performs a homology search on double-stranded DNA and catalyzes the ATP-dependent formation of a joint molecule (*D-loop*, displacement loop) with homologous sequences. Synapsis results in the formation of intermediates, so-called *Holliday junctions*. Postsynapsis involves movement of the crossing-over regions (*branch migration*) and finally resolution of the DNA strands by specific cutting reactions.

4.1.1.1 Presynapsis

General Considerations Initiation of recombination in *E. coli* (and possibly in all organisms) requires the formation of ssDNA. Single strands are generated by the combined helicase–nuclease activities of the RecBCD multifunctional enzyme complex (Table 4.1). The presynaptic step carried out by the heterotrimeric RecBCD enzyme complex fullfills two tasks:

1. It protects the cell from linear foreign dsDNA such as the genomes of bacteriophages λ and T4 upon entering the cell if it is not accompanied by a protective mechanism.
2. It helps the cells to recover from double-stranded breaks in its chromosome during genomic replication.

The RecBCD complex faces a dilemma during its interaction with linear dsDNA: whether or not to degrade it completely. Invading foreign DNA should be completely destroyed, whereas an *E. coli* genome that has suffered a double-stranded break should be preserved via recombinational repair. Nature has an ingenious solution for the cell, in the form of an 8-bp sequence called Chi (for *c*rossover *h*otspot *i*nstigator; also called χ; 5′-GCTGGTGG-3′). Chi appears at a high frequency

in the *E. coli* chromosome, marking it for protection. The absence of Chi in a double-stranded DNA (e.g. in phage λ) defaults it to complete degradation. Chi, originally discovered as a recombination hot spot, stimulates recombination in its vicinity 5- to 10-fold. Null mutations in *recB* or *recC* render the cell recombination deficient and sensitive to DNA damage. RecBCD needs linear DNA, but with a few exceptions, all native molecules are circular. What processes convert circular into linear DNA molecules?

1. Most double-stranded breaks occur at the replication fork.
2. Double-stranded breaks can be induced by ionizing radiation.

The Mechanism of Presynapsis The presynapsis starts with the binding of the heterotrimeric 330-kDa RecBCD enzyme complex to linear ends of dsDNA (Fig. 4.1). RecB and RecD have the ability to unwind DNA (they are helicases) and these two motors move in the same direction along a DNA double helix, but on opposite strands. While RecB progresses in the 3′ to 5′ direction, RecD moves in the 5′ to 3′ direction. Both proteins are endonuleases as well, where RecB cleaves frequently the DNA strand ending with the 3′ end and RecD cuts less frequently the complementary strand. The RecC protein fullfills two tasks: First, it keeps RecB and RecD together in one complex, and second, it recognizes the Chi sequence as soon as it becomes single-stranded. When the RecBCD encounters a

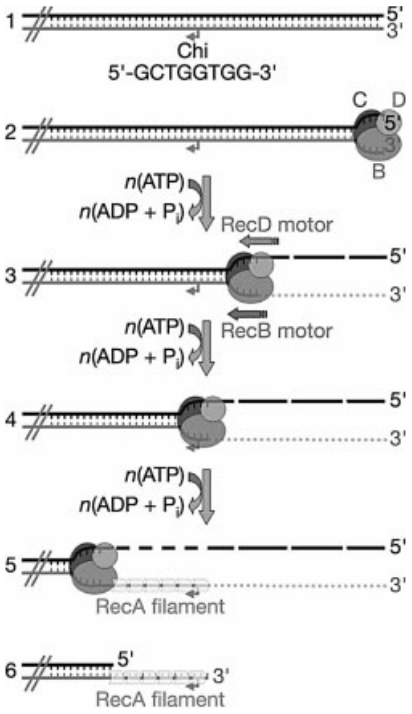


Fig. 4.1 The presynaptic step. Stage 1 shows linear, double-strand DNA with a Chi sequence located in the middle. In stage 2, a RecBCD complex is bound to one end. The complex starts to unwind the DNA double helix with the aid of the motor proteins RecD and RecB consuming ATP. Both proteins are held together by RecC, and degrade both strands. RecB cuts frequently, RecD occasionally (stage 3). When the complex arrives at a correctly positioned Chi site which is recognized by RecC, it pauses and changes its nucleolytic activity (stage 4). During stage 5, the complex continues unwinding the DNA duplex, leaving the 3′ ending single strand intact. RecA is loaded on the intact DNA strand forming a RecA filament. In stage 6, the RecBCD complex dissociates from its substrate. M.R. Singleton **2004**, *Nature* 432, 187; Fig. 1.

properly oriented Chi sequence, the complex is switched from the *destructive mode* to a *recombinogenic mode*. The Chi octamer elicits biochemical changes in RecBCD enzyme: (a) the enzyme briefly pauses at Chi before unwinding resumes, (b) the nuclease activity is reduced and its polarity is switched. The final cleavage event on the 3′ tail occurs at or within a few bases to the 3′ side of Chi; and the 3′ tail is then protected from further digestion. Finally, the RecBCD complex loads the RecA protein onto the 3′ tail, forming the RecA filament which is used to initiate the homologous pairing phase of genetic recombination. The RecBCD enzyme is the most potent exonuclease in *E. coli* (ExoV) and can degrade more than 30 kb of DNA per initial binding event, with the helicase unwinding DNA at rates of the order of 1000 bp s^{-1}.

Chi is a ssDNA cofactor that acts as an effector to change the antirecombinogenic ExoV nuclease activity to a recombinase. The Chi sequence is overrepresented, i.e. it is much more frequent than expected for a random sequence of the same size. It has been suggested that the canonical 8-base Chi sequence in *E. coli* usually exists within approximately 1-kb GT-rich 'recombination islands', which is supported by the finding that RecA binds preferentially to GT-rich DNA. The presence of Chi sequences on the genome has not a random occurrence. It is notable that 75% of the Chi sites on the *E. coli* chromosome are cooriented with respect to the direction of DNA replication, in agreement with the observation that most double-strand breaks occur at the replication fork. It is suggested that this orientation of Chi on the chromosome is related to its role in stimulating DNA repair. Several bacteria have Chi-like sites as short sequences that are able to attenuate the bacterial exonuclease activity. Its sequences have been determined as 5′-GCGCGTG-3′ in *L. lactis*, 5′-AGCGG-3′ in *B. subtilis* and 5′-GNTGGTGG-3′ and 5′-G(G/C)TGGAGG-3′ in *H. influenzae*.

4.1.1.2 **Synapsis**

Synapsis between homologous DNA is catalyzed by the ubiquitous family of RecA-like proteins. The members of this family are conserved among Eubacteria, Eucarya and Archaea. The most widely studied recombinase is the 38-kDa RecA protein of *E. coli*. Under normal growth conditions, there are about 8000 to 10000 monomers of RecA per cell, and this number increases to over 70000 upon treatment with DNA-damaging agents inducing the SOS response. Besides playing a pivotal role in homologous recombination, RecA is required for induction of the SOS pathway of DNA repair and mutagenesis, where it acts as a coprotease (see Section 5.7.9) and exerts an important role in the repair of stalled replication forks. On binding ATP, RecA polymerizes on single-stranded DNA with a site size of three bases per protein and forms a right-handed helix (Fig. 4.2). Filament assembly proceeds with 5′→3′ polarity. The DNA within the filament is stretched from 10.5 bp per turn (B-form DNA) to 18.6 bp per turn as RecA imposes its own helicity on the DNA. RecA is also a DNA-dependent ATPase: it hydrolyzes ATP when bound to DNA and switches between a high-affinity state, RecA-DNA-ATP, and a low-affinity state, RecA-DNA-ADP. It has generally been considered that

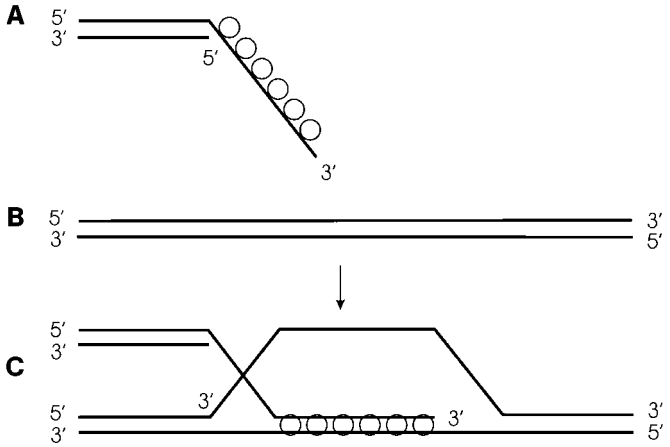


Fig. 4.2 The synaptic step. The RecBCD enzyme leaves as ssDNA onto which RecA can load (A) and screen dsDNA (B) for a homologous region. It then anneals to the complementary strand (C), simultaneously displacing the second strand.

RecA binding to single-stranded DNA is not specific, but RecA has a preference affinity for DNA sequences rich in GT, as already mentioned. Single-stranded DNA forms a complex with RecA at site I, known as a presynaptic filament. Double-stranded DNA then binds at a second site, site II. If the DNA sequence at site I and II are complementary, synapsis can occur, leading to strand exchange and thereby forming a Holliday junction. Strand exchange is unidirectional (5′→3′ with respect to the single strand within the filament) and relatively slow (3–10 bp s^{-1}). As strand exchange proceeds into a double-stranded region of the first DNA molecule, a Holliday intermediate is formed. Of all steps in DNA strand exchange, the most mysterious is the manner by which DNA sequence homology is recognized. How the RecA nucleoprotein filament makes the normally stable dsDNA receptive for DNA strand invasion remains a key question. It should be mentioned that the protein RecX inhibits RecA recombinase and coprotease activities by direct protein–protein interaction to attenuate the deleterious effects of

recA overexpression.The direct association of RecX with RecA severely impedes ATP hydrolysis and the generation of heteroduplex DNA. The *recX* gene occurs in many bacterial species and is normally located downstream of *recA* or overlapping *recA* or occasionally elsewhere in the chromosome.

4.1.1.3 **Postsynapsis: the RuvABC Resolvasome**

During the late stages of recombination, Holliday junction intermediates are processed into mature recombinants by the three proteins RuvA, RuvB and RuvC. The RuvA tetramer is a junction-specific binding protein that interacts directly with the hexameric RuvB and loads two hexameric RuvB helicase rings on opposite arms of the Holliday junction (Fig. 4.3). The RuvA monomer (22 kDa) con-

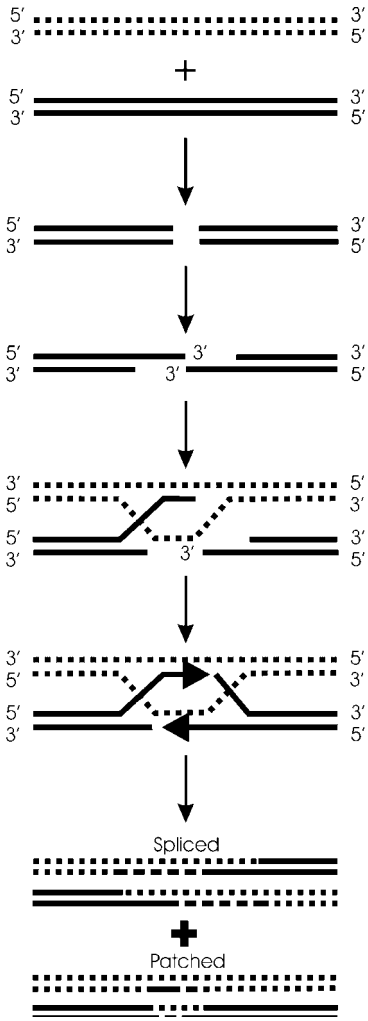


Fig. 4.3 The postsynaptic step. A double-stranded break has occurred in one of the two homologous DNA molecules, and the break will be converted to single-stranded DNA with 3′ overhangs by the RecBCD complex. RecA will bind to the single-stranded regions and invade the intact double-stranded DNA molecule creating a Holliday junction which will be recognized by the Ru-vABC complex. Depending on where resolution will occur, either the two DNA molecules will be spliced together or patched molecules will result.

sists of three domains: domains I and II are involved in tetramer formation and Holliday junction recognition, respectively. Domain III is highly mobile, is connected with domain II via a flexible loop and is involved in a specific interaction with RuvB. RuvB is thought to constitute the motor that drives branch migration using energy derived from ATP hydrolysis. The third component, a dimer of RuvC, binds to the branch point on the opposite face of the Holliday junction to RuvA and cleaves the junction by introducing symmetrical nicks in the continous (noncrossover) strands. Cleavage occurs preferentially at the consensus sequence 5′-(A/T)TT↓(G/C), where ↓ represents a cleavage site close to the crossover point. Following cleavage, the resolution process is completed by DNA ligase, which rejoins the 5′ phosphate and 3′ OH termini in the nicked duplex products.

4.1.2
The RecF Pathway of Homologous Recombination

The recombination-deficient phenotype of *recBC* mutants is suppressed in cells mutant in both *sbcB* (suppression of *recBC*) and *sbcCD*. While *sbcB* encodes the 3′ to 5′ ssDNA exonuclease I, *sbcCD* encodes the dsDNA exonuclease SbcCD. In the absence of both inhibitory nucleases, recombination intermediates are preserved. The linear DNA is unwound by the RecQ protein (helicase), followed by digestion of the 5′ end by RecJ (5′→3′ ss exonuclease), leaving 3′-tailed ssDNA coated with SSB (Fig. 4.4). Next, SSB is replaced by RecA which is loaded by RecF, RecO and RecR proteins, by a mechanism poorly understood. The current view is that the RecOR complex stimulates RecA-ssDNA filament formation, while the RecFR complex prevents its extension when it reaches dsDNA. In summary, the RecF

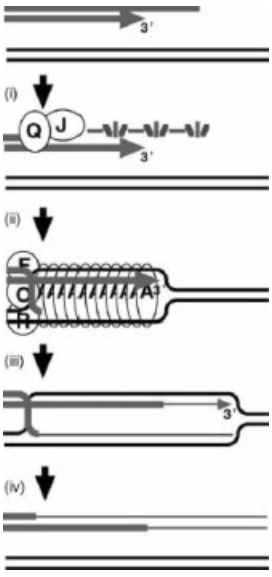


Fig. 4.4 The RecF pathway of homologous recombination. The RecQ helicase binds to linear ends and unwinds the helix. (i) The 5′→3′ exonuclease RecJ removes the upper strand, while RecF promotes binding of RecA to the single-stranded lower strand ending in 3′. Next, the RecA-coated strand invades homologous dsDNA (ii); and replication is initiated from the free 3′ ends (iii), followed by cleavage of the crossover products (iv). J. Courcelle et al. **2004**, *Res. Microbiol.* 155, 231–237; Fig. 3B.

pathway requires the activities of several separate proteins to form the presynaptic complex.

Several studies have revealed a link between DNA replication and the RecF pathway, including phage λ recombination, long-patch excision repair and certain forms of aberrant replication, such as plasmid linear multimer formation. While *recF* and *recR* mutants have relatively subtle phenotypes with respect to recombination, their UV sensitivities are comparatively dramatic. These observations suggest that recombination is not the primary function of the RecF pathway, but resumption of DNA replication from existing replication forks requires both the *recF* and *recR* genes. Both proteins are required to reassemble a replication holoenzyme at the site of a DNA replication fork, suggesting that the UV hypersensitivity of *recF* cells is due to an inability to resume replication at disrupted replication forks rather than to a defect in recombination.

4.1.3
Additional Homologous Recombination Functions

The *recE* gene was discovered by isolating another class of suppressors of *recBC* mutations, called *sbcA*, that restored recombination. The *sbcA* mutations were later shown to activate a normally repressed recombination function of the defective prophage Rac. It is assumed that the *sbcA* mutations inactivate a repressor gene that normally prevents transcription of the *recE* and *recT* genes, whose products can substitute for the RecBCD nuclease and for RecA in recombination. When *recE* and *recT* are expressed, many types of recombination events occur in the absence of RecA, including plasmid recombination, double-strand break and gap repair. In addition, the RecET system can promote recombination between short homologous sequences such as occurs in PCR construct-mediated gene-targeting events apparently not promoted by the RecA-dependent recombination pathway. The RecE protein, also called exonuclease VIII, is an ATP-independent exonuclease that preferentially degrades linear dsDNA in the 5′→3′ direction and also degrades ssDNA at low rates. The RecT protein, in the absence of ATP, promotes homologous pairing and strand exchange between ssDNA and dsDNA, thereby replacing RecA. Also, the *ruvAB* genes can be substituted by another gene designated *recG*. If one of the two *ruv* genes is inactive, the *recG* gene product is still available to recognize Holliday junctions. The product of *recG* is a 76-kDa structure-specific DNA helicase, which can catalyze the branch migration of Holliday junctions. Resolution of the Holliday junctions, in the absence of *ruvC*, can occur by the *rusA* gene product. This normally silent gene is located on another defective prophage, called DLP12, and its transcription can be activated by insertion of IS*2* or IS*10* elements upstream of the *rusA* coding region. The 14-kDa RusA protein is a homodimeric Holliday junction-specific endonuclease. It is sequence-specific like the RuvC resolvase, although it prefers to cut 5′ of a CC dinucleotide.

4.1.4 Recovery of Replication at a Blocking DNA Lesion

One potential source of genome instability is collision of the DNA replication machinery with unpaired lesions in the template that block progression of the replication fork. These occur, for instance, after irradiation of cells with near-UV light. Such an event creates two major problems. First, the original block must be repaired or removed. Second, replication must be restarted with the highest fidelity. The price of failure is genetic damage ranging from single-point mutations to large-scale chromosomal rearrangements. How do cells restart replication in an error-free manner? Five possibilities have been described:

- replication restart by Holliday junction formation,
- replication fork regression and processing,
- rescue of damaged replication forks by homologous recombination,
- translesion bypass synthesis,
- daughter-strand gap repair (postreplication recombination repair).

How often do replication fork problems arise? Based on experimental data, it has been calculated that 18% of the cells require replication-fork reloading during a single round of chromosome duplication in the absence of any exogenous DNA-damaging agent.

Replication Restart by Holliday Junction Formation

Stalled forks might unwind so that the parental strands reanneal and the nascent daughter strands anneal to from a Holliday junction. This may occur spontaneously via release of positive supercoiling ahead of the fork, but is more likely catalyzed either by RecG helicase or via the strand exchange activity of RecA, aided perhaps by the RecF, RecO and RecR proteins. Once formed, the Holliday junction may be driven further from the lesion by the RuvAB branch migration complex. Such reactions would require the replisome complex to have dissociated, which raises the question of how DNA synthesis might subsequently resume. Furthermore, the lesion has to be repaired or bypassed for this renewed synthesis to continue. Fork reversal and Holliday junction formation provide possible solutions to these problems.

Replication Fork Regression and Processing

DNA lesions blocking replication fork progression induce regression of the fork, involving reannealing of the parental strands and the two daughter strands (Fig. 4.5). This intermediate is maintained by the proteins RecA, RecF, RecO and RecR and allows repair enzymes to gain access to the DNA lesion. In the absence of the Rec proteins, the intermediate is degraded by the combined action of RecQ and RecJ. Once the blocking lesion has been repaired, processive replication resumes.

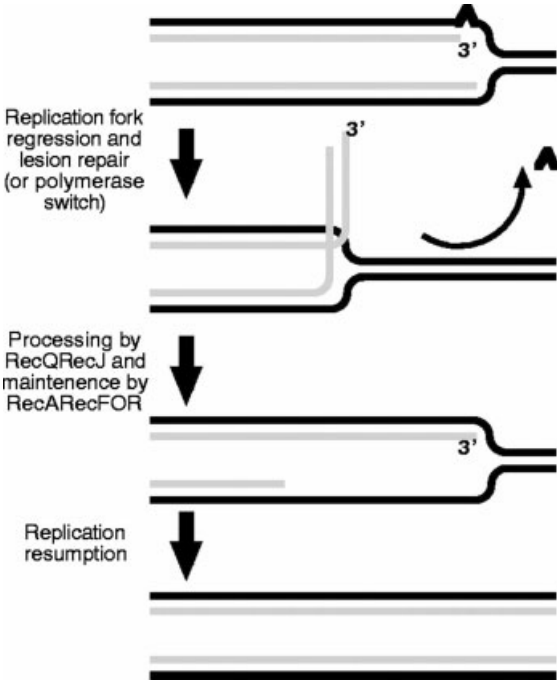


Fig. 4.5 Model for the structural intermediates associated with the recovery of replication blocked at a DNA lesion (∧). J. Courcelle **2003**, *Science* 299, 1064; Fig. 4.

Rescue of Damaged Replication Forks by Homologous Recombination

When the replication fork encounters a nick in either the leading or the lagging strand template, this generates a double-strand break in one of the sister chromosomes, which can be repaired by homologous recombination (Fig. 4.6). Upon arriving at the nick (shown for the leading strand in Fig. 4.6), the replisome falls apart (steps i, ii). The gap left at the lagging strand is filled in by a so far unknown DNA polymerase and the nick sealed by DNA ligase (iii). The broken chromosome is recognized by the RecBCD enzyme degrading preferentially the 3′ terminal strand, until it encounters a Chi site where degradation of the 5′-terminated strand starts (iii). Next, RecA binds to the 3′-terminated strand and invades the sister duplex (iv), forming a displacement (D)-loop (v). The D-loop is recognized by PriA which organizes assembly of the replication restart primosome, which is composed of at least PriB, DnaT, DnaB, DnaC and DnaG. The major function of the primosome is to load DnaB onto the lagging strand template.

Translesion Synthesis

The high fidelity of Pol III and its inability to use damaged DNA templates creates the need for *trans*lesion *s*ynthesis (TLS). Three different nonessential DNA polymerases have been identified as being able to carry out translesion synthesis: Pol II, Pol IV and Pol V. Among these, Pol II (dissociated Pol III) is a high-fidelity

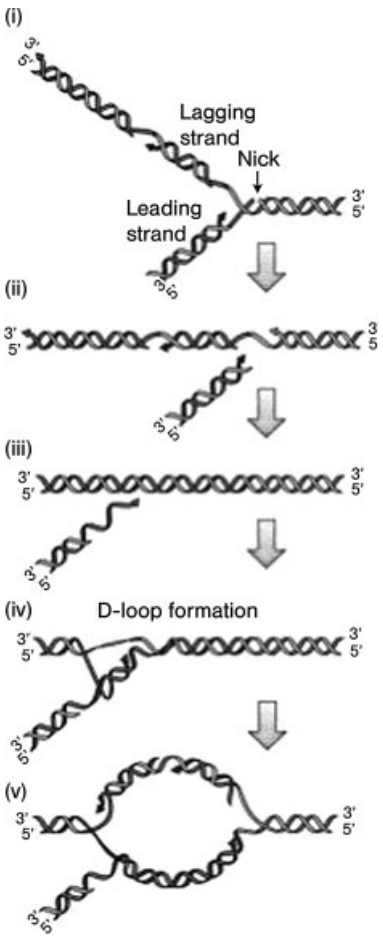


Fig. 4.6 Sequence of events leading to the rescue of a damaged replication fork. See text for details. K.J. Marians **2000**, *Trends Biochem. Sci.* 25, 185–189; Fig. 1.

polymerase that uses error-free replication restart at UV-induced lesions by a mechanism called *lesion avoidance replication*. Arrived at a DNA lesion, Pol III dissociates and is replaced by Pol II. If Pol III encounters a DNA lesion, it can be replaced by Pol II which switches template and continues replication using the complementary daughter strand before it resumes at the original template. Then, Pol III substitutes for Pol II again, switching template to continue replication on the newly synthesized daughter strand complementary to the original parental strand to circumvent the lesion. Thereafter, it switches template again, continues on the original parental strand and is replaced by Pol III.

The two other polymerases, Pol IV and Pol V are error-prone enzymes. Pol IV encoded by *dinB* is normally abundant but is further induced around ten-fold by the SOS response. Pol IV has been shown to be involved in the mutagenesis of undamaged bacteriophage λ, when grown on a UV-irradiated host. Pol IV is incapable of bypassing UV-induced pyrimidine dimers, but it can bypass some

bulky lesions; and it is also able to extend a misaligned 3′ primer with a bulged template, leading to a replication product one nucleotide shorter than expected. This ability is consistent with the propensity of Pol IV-dependent –1 frameshift events observed *in vivo*. It has been shown that Pol IV is responsible for the majority (~85%) of *lacZ* –1 frameshift mutations. But, *in vivo*, most translesion DNA synthesis is dependent on the damage-inducible heterotrimeric complex of $UmuD'_2C$, where *umuC* codes for DNA polymerase V (Pol V). This system is very powerful in translesion synthesis, but at the expense of mutations. *E. coli* strains with mutations in either *umuC* or *umuD* do not exhibit the dramatically elevated mutation rates seen when wild-type cells are exposed to various mutagens. Therefore, the Umu-dependent pathway of translesion DNA synthesis is *error prone*. The reconstitution of translesion replication with purified components revealed that lesions are bypassed by Pol V in the presence of UmuD′, RecA and SSB. Pol V can bypass a synthetic abasic (apurine/apyrimidine) site, a *cis-syn* cyclobutyl TT dimer, and a 6-4 TT photoproduct, usually in a mutagenic manner. It is also highly mutagenic when replicating undamaged DNA and preferentially forms transversion mutations. The proteins involved in copying blocking template lesions are Pol V, RecA, SSB, the β-sliding clamp and the γ-clamp loading complex; and all these different proteins together have been designated ''Pol V Mut'', where ''Mut'' refers to the *mutasome*.

The different steps of translesion synthesis (TLS) by Pol V Mut are presented in Fig. 4.7: (a) Pol III holoenzyme encounters a template lesion (X) where it stops, while replicative DnaB helicase continues to unwind the DNA downstream of the lesion; (b) Pol III core dissociates from the 3′ primer end and an activated RecA nucleoprotein filament is assembled in a 5′ to 3′ direction on the unwound ssDNA; (c) finally, the RecA nucleofilament reaches the site of DNA damage; (d) then, Pol V binds to the 3′ end of the primer through interaction with the β sliding clamp and this reaction is stimulated by RecA and SSB; (e) Pol V and SSB dissociate RecA from the DNA template in a 3′ to 5′ direction and polymerizes across the lesion; (f) after TLS, Pol V dissociates from the DNA and is replaced by Pol III.

Lesion bypass by Pol V requires assembly of a RecA nucleoprotein filament, which targets the polymerase to the primer terminus. SSB facilitates the formation of the RecA nucleoprotein filament and the unloading of RecA molecules, which is a necessity for Pol V replication activity at both the initiation and elongation and the bypass stages. UmuD′ may serve as a bridge between Pol V and the RecA nucleoprotein, since it interacts with both proteins, while direct interaction between UmuC and RecA has not been observed.

Daughter-strand Gap Repair (Postreplication Recombination Repair)

If the replication fork encounters a DNA lesion, it can skip over the damage and restart downstream leaving a single-stranded gap (Fig. 4.8). The replication fork skips over DNA lesions, leaving single-stranded gaps. These are protected by RecA, which subsequently transfers and anneals the complementary strand to the gap. After resolution of the crossover junctions, the gaps in the parental strands

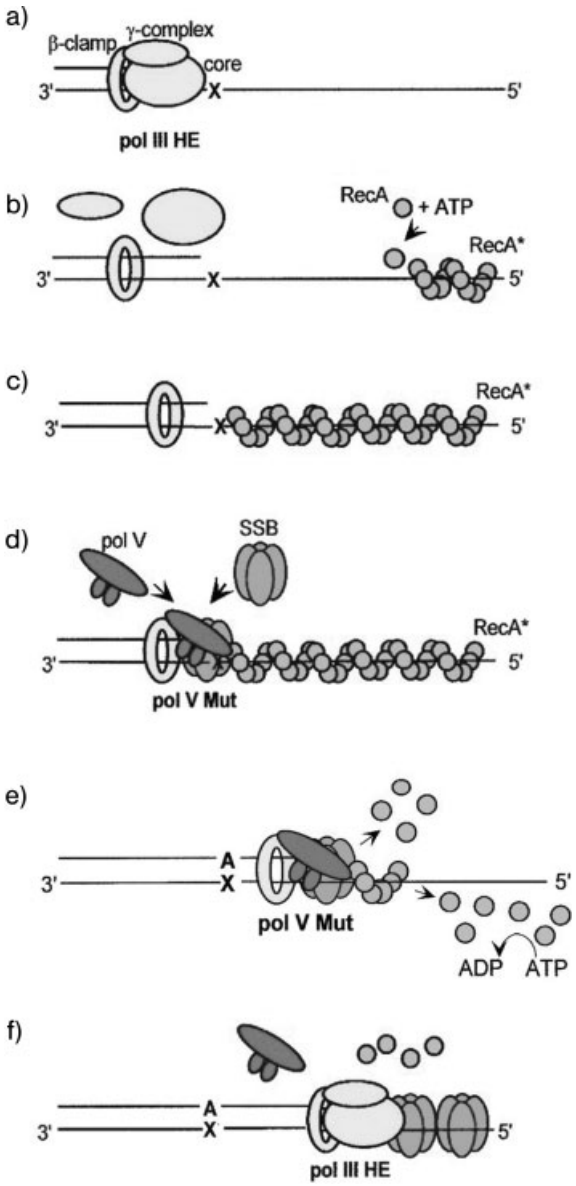


Fig. 4.7 Model describing DNA polymerase V mutasome-catalyzed TLS. See text for detailed explanation. P. Pham **2001**, *Proc. Natl. Acad. Sci. USA* 98, 8350; Fig. 2.

are filled by DNA synthesis. Next, one of the different DNA repair systems can remove the damage before another replication fork arrives at the DNA lesion. This is recognized by RecA binding to the gap and transferring the complementary parental strand to the gap, followed by annealing. Next, the crossover junctions are resolved and the gap is filled-in by DNA synthesis. In principle, the gap is transferred opposite to the intact DNA strand, leaving time for repair of the DNA lesion until the next replication fork arrives.

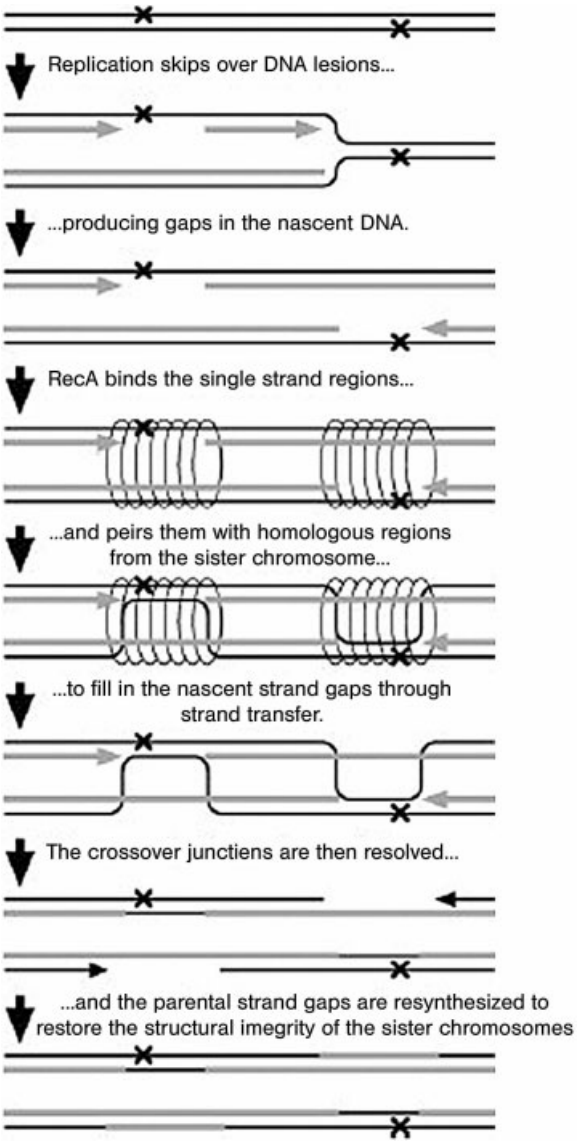

Replication skips over DNA lesions...
...producing gaps in the nascent DNA.
RecA binds the single strand regions...
...and peirs them with homologous regions from the sister chromosome...
...to fill in the nascent strand gaps through strand transfer.
The crossover junctiens are then resolved...
...and the parental strand gaps are resynthesized to restore the structural imegrity of the sister chromosomes

Fig. 4.8 Model for daughter-strand gap repair. See text for detailed description. J. Courcelle **2003**, *Annu. Rev. Genet.* 37, 611–646; Fig. 2.

4.2
Sequence-specific Recombination

Sequence- or site-specific recombination differs from general homologous recombination in a number of respects. General homologous recombination events require substrate DNAs that share considerably homology, with exchange occurring anywhere within the region of homology. The frequency of recombination increases with increasing length of the homologous regions. In addition, proteins of

the general recombination pathway, especially RecA, are required for general homologous recombination. In contrast, recombining sequences in site-specific reactions are usually short; and the reaction occurs at a single specific site within the recombining sequence and involves the formation of a Holliday junction intermediate. Furthermore, sequence-specific recombination usually requires the presence of a specific recombinase and one or more cofactors and is independent of the general recombination pathway. It is a conservative process in which all DNA strands that were broken are rejoined without ATP utilization or DNA synthesis. Site-specific recombinases are divided into two families, the *resolvase/invertase family* and the *integrase family*. Site-specific recombinases of both families are utilized in various recombination systems, including inversion, integration, excision and resolution of cointegrates. Different site-specific recombination systems carry a varying complexity of recombination sites. The sequence where actual DNA crossover occurs are called *recombination core sites*. Core sites are ~30 bp long and normally consist of a pair of inverted repeats that act as binding sites for the recombinase, separated by a spacer of 6–8 bp. Two site-specific recombinase molecules bind head-to-head to the repeat sequence. The spacer sequence of both recombining core sites must be identical and determines the orientation of each core site. The sequence homology of the spacer region is now thought to be gauged during the annealing step prior to strand-joining without branch migration.

The resolvase/invertase family of recombinases promotes the resolution of cointegrate intermediates of Tn*3* group transposons and the inversion of H, G, C and P segments by binding to the recombination core site termed *res*. The proteins are approximately 200 amino acid residues in length and mediate a concerted double-strand cleavage and rejoining reaction, which produces intermediates with 2-bp staggered cuts and 3′ protruding ends. An active center serine residue of the recombinase molecule forms a phosphodiester bond with the 5′ end of the recombining DNA strand. The crystal structure of γδ resolvase has been elucidated. A dimer of resolvase molecules binds to a DNA duplex, inducing a 60° DNA kink at the center of the *res* site. The N-terminal two-thirds of resolvase form a globular domain with an active center serine residue, whereas the C-terminal one-third forms an arm and a helix-turn-helix DNA-binding domain. Over 140 members of the resolvase/invertase family have been characterized.

The integrase family includes about 200 highly diversified members, such as λ integrase, Cre, Flp and XerC-XerD. The integrase nucleophile is a tyrosine located at the C-terminal end of the protein. This tyrosine is responsible for the cleavage and forms a covalent intermediate by esterification of a DNA 3′ phosphoryl group. The joining reaction is mediated by nucleophilic attack of the 5′ hydroxyl group from the same strand or another cleaved strand. The integrase family definition is based on identification of conserved residues (RHRY) found in two boxes (I, II) located in the carboxyl half of the protein. Only the histidine is not invariant, being present in nearly 95% of the members.

4.2.1 Integration/Excision Systems

The best known integration/excision systems are those encoded by many temperent phages such as phage λ. After infection of a sensitive *E. coli* cell, the λ DNA forms a circle by pairing between the *cos* sequences (12-nucleotide single-stranded ends complementary to each other). The Int protein can then catalyze integration of the circular λ DNA into the bacterial chromosome by promoting site-specific recombination between two attachment sites. One, *attP* (for *att*achment *p*hage), is located on the phage DNA while the second, *attB* (*att*achment *b*acteria), lies between the galactose and biotin operons on the *E. coli* chromosome. Both attachment sites have a common core sequence, O, of only 15 bp (5′-GCTTT[TTTATAC]TAA-3′) flanked by two dissimilar DNA sequences, B and B′, in *attB*, and P and P′ in *attP*. Recombination always occurs within the 7-bp sequence shown in squared brackets. Besides the λ Int protein, two *E. coli* proteins are involved in the recombination process, IHF and Fis, where the former is essential and the latter stimulates the recombination process. The integration process is reversible, but needs a fourth protein, the Xis protein, also encoded by the λ DNA. The Xis protein (excisionase) is required to recognize the hybrid attachment sites *attBOP* and *attPOB*. Besides this *primary attachment* site, there are so-called *secondary attachment* sites, where λ DNA can integrate at a low frequency (<1%). These secondary attachment sites deviate from the primary site in several nucleotides (see Table 4.2). Sometimes, excision does not involve the hybrid attachment sites but a different sequence, leading to the production of specialized transducing phages (see Section 10.3.2).

This table shows the DNA sequence attachment site of phage λ (*attP*) and primary attachment site of *E. coli* first (*attB*), followed by five secondary attachment sites. Deviations from the primary site are displayed in lowercase letters.

Table 4.2 Primary and secondary attachment sites in the chromosome of *E. coli* allowing integration of λ DNA.

Site	*Core DNA sequence*		
attP	TCAGCTTT	TTTATAC	TAAGTTGG
attB	CCTGCTTT	TTTATAC	TAACTTGA
proB	tgcGCTAA	TTTATAC	gAggcTac
trpC	gCgtaaTg	TTTATAa	atggcgGc
galT	cgcctTTg	TTTtcAa	aAAccTGc
thrA	cggGcTTT	TTTcTgt	gtttcctg
rrnB	ttgGcTaT	TTTAcca	cgACTgtc

4.2.2 Resolvase Systems

Resolvase systems are involved in the resolution of replicon multimers. These are either formed by homologous recombination and lead to plasmid multimers or chromosome dimers or are formed by illegitimate recombination and lead to cointegrate structures. We will consider here three examples: resolution of plasmid multimers, chromosome dimers and a cointegrate structure formed by the Tn*3* transposon.

Resolution of Plasmid Multimers

Stable maintenance of bacterial plasmids requires efficient replication control and distribution to daughter cells. High-copy-number plasmids are distributed randomly at cell division and the probability of forming a plasmid-free daughter cell (p_0) is given by $p_0 = 2^{(1-n)}$, where n is the plasmid copy number of the dividing cell. Any factor which reduces the plasmid copy number increases the frequency of plasmid loss. Multimer formation by homologous recombination or by rolling-circle replication is an important cause of multicopy instability, as multimer-containing cells contain fewer independent plasmids. The multicopy plasmid ColE1 counteracts the effects of multimerization through the activity of the Xer sequence-specific recombination system. This system consists of the 240-bp recombination site *cer* (Col*E*1 *r*esolution) located on the plasmid and four proteins (XerC, XerD, ArgR, PepA) encoded by the *E. coli* chromosome (Fig. 4.9). The recombination reaction is catalyzed by XerC and XerD, two members of the integrase class of site-specific recombinases. The *ar*ginine *r*epressor (ArgR) and amino*pep*tidase *A* (PepA) are accessory proteins which are not directly involved in the strand exchange reaction but are absolutely required for recombination at the *cer* site *in vivo* and *in vitro*. ArgR is an L-arginine-dependent DNA-binding protein that acts as a transcriptional repressor of the arginine regulon: it binds as a 102-kDa homo-hexamer to 18-bp ARG boxes which are usually present in two copies separated by 3 bp, in the promoter regions of the genes for arginine biosynthesis. In the present case, ArgR binds to a single ARG box within *cer*, ~110 bp from the point of strand exchange, and induces a bend of ~65°. It seems likely that, during

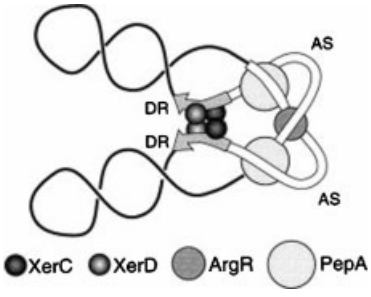


Fig. 4.9 Resolution of ColE1-dimers. The XerCD heterotetramer binds to the two *cer* sequences (arrows) to align them (synpasis). Formation of synapsis is aided by the two DNA-binding proteins ArgR and PepA. B. Hallet, D.J. Sherratt **1997**, *FEMS Microbiol. Rev.* 21, 157–178; Fig. 7c.

recombination at *cer*, a single ArgR hexamer binds to one ARG box from each participating *cer* site, helping to synapse two *cer* sites and/or introducing a structurally important bend within the accessory sequences. PepA is one of the major aminopeptidases and is thought to be important for the metabolism of peptides supplied exogenously or produced by protein degradation within the cell. PepA is a hexamer in solution consisting of six identical 55-kDa subunits. In addition, PepA has a DNA-binding activity (binds specifically to sequences overlapping the pyrimidine-regulated P1 promoter of *carAB*) and this activity is important for its function in Xer recombination. PepA binds specifically to the accessory sequences of *cer* and cooperates with ArgR to make an interwrapped complex with the accessory sequences of two *cer* sites. Thereby, PepA plays a major role in defining the interwrapped structure of this complex. The *cer* site consists of a ~30 bp core sequence, containing an 11-bp XerC-binding and an 11-bp XerD-binding site separated by an asymmetric central region of 6–8 bp, and ~180 bp of accessory sequences adjacent to the core. ArgR and PepA bind to the accessory sequences of *cer* and form a complex in which two recombination sites are interwrapped in a right-handed fashion. Xer recombination at *cer* is exclusively intramolecular, resolving but not creating plasmid multimers. ArgR and PepA appear to be responsible for ensuring this resolution selectively during recombination at *cer*. A similar resolution sequence has been identified on the medium-copy-number plasmid pSC101 and termed *psi* (*p*SC101 *s*tabilized *i*nheritance). Here, the accessory sequences of *psi* contain binding sites for PepA and the anaerobic repressor protein ArcA.

A more severe problem due to the formation of dimers occurs with low-copy-number plasmids being present only in one copy per chromosome, such as the F factor or the P1 prophage. If dimerization occurs shortly before segregation into the two daughter cells, the dimer can segregate only into one of the two daughter cells and the plasmid-less cells then overgrow the others. To prevent loss of the plasmid, efficient site-specific resolution systems resolve the plasmid dimers. In both cases, they consist, in contrast to the XerCD system, of only one recombinase and two recombination sites. In the case of the P1 prophage, the Cre (for *c*yclization *re*combination) recombinase resolves dimers at the two *loxP* sites (for *lo*cus of *X*-over, *p*hage; 24 bp). The latter system is so efficient that it is used in many genetic systems, including eukaryotic ones such as *Drosophila*, plants and mouse. In these systems, genetic information to be deleted is sandwiched between *loxP* sequences in direct orientation and expression of the Cre-recombinase is turned on when appropriate.

Resolution of a Chromosome Dimer

Crossing over by homologous recombination can lead to dimerization of circular chromosomes, thereby compromising their segregation. Consistent with this, eubacteria with circular chromosomes make use of the specialized Xer site-specific recombination system to convert chromosome dimers to monomers (see Section 3.1.1.5). In *E. coli*, the two Xer site-specific recombinases XerC and XerD act on a 28-bp chromosomal recombination site, *dif* (*d*eletion-*i*nduced *f*ilamentation), lo-

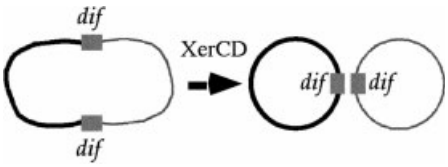


Fig. 4.10 Resolution of a chromosomal dimer. The recombinases XerCD bind to the recombination sites *dif*, bring these two sites together and catalyze site-specific recombination to resolve a dimer into two monomers. K.P. Lemon **2001**, *Genes Dev.* 15, 2031; Fig. 4B.

cated in the replication terminus of the chromosome, to ensure faithful segregation of newly replicated chromosomes to daughter cells at cell division. Recombination at *dif* (consisting of 11-bp XerC and XerD binding sequences, separated by a 6-bp spacer) does not require ArgR and PepA. When a chromosome dimer is present, the XerCD recombinase complex resolves the dimer into two monomers (Fig. 4.10). The Xer recombination reaction is highly regulated to ensure proper chromosome dimer resolution without risking the reverse reaction, which would create dimers from monomers. This is achieved by placing the Xer recombination system under the control of the cell division protein FtsK and through the location of the recombination site *dif* at a specific position in the replication terminus region of the chromosome that is actively maintained at midcell during most of the cell cycle. FtsK is a bifunctional protein whose N-terminal part is required for cell division and is responsible for its association with the division septum (see Section 3.3). The C-terminal part of FtsK is dispensable for cell division but is strictly required for chromosome dimer resolution, suggesting a functional interaction between the XerCD-*dif* complex and the division septum. FtsK acts to align the *dif* regions at the cell center, thus allowing a productive synapse between the *dif* sites. In *B. subtilis*, the *dif* site and the XerCD homologs, CodV and RipX, have been identified and shown to function in dimer resolution, too.

It should be mentioned that the ~6.9-kb CTXΦ *V. cholerae* filamentous phage integrates into the host chromosome by making use of the XerCD recombinases and the integration site overlaps with the *dif* site of the larger of the two chromosomes.

Resolution of Cointegrates Formed During Tn*3* Transposition

Mobile elements transposing by the replicative mechanism from one replicon to another (see below) form an intermediate called the *cointegrate*. In the cointegrate, the donor and target DNAs have become fused and encode two copies of the mobile element which separate the two replicons. Normally, these cointegrates are rapidly resolved in the two replicons; and this process has been studied in detail using the transposon Tn*3* and Tn*1000* (also called γδ; present on the F factor). Cointegrate resolution occurs by sequence-specific recombination and needs two *cis*-elements called *res* sequences (90–140 bp long) joined in appropriate head-to-tail orientation on the same superhelical DNA molecule. The typical *res* sequence contains three binding sites (I, II, III) for a resolvase dimer. Site I contains the site of DNA breakage and strand exchange, while the two other sites, II and III, are crucial for both synapsis of the two *res* sequences and to ensure directional specificity.

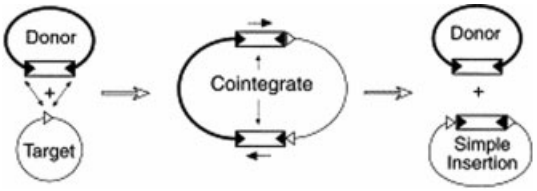


Fig. 4.11 Resolution of a cointegrate, an intermediate of Tn*3* transposition. During replicative transposition, a mobile element such as Tn*3* forms an unstable cointegrate structure which is subsequently resolved by site-specific recombination through the resolvase enzyme. B. Hallet, D.J. Sherratt **1997**, *FEMS Microbiol. Rev*. 21, 157–178; Fig. 1.

Most resolvases are small proteins (180–210 amino acid residues). The Tn*1000* resolvase consists of two globular domains joined by a long extended linker. The N-terminal domain of about 120 residues contains the catalytic site and is responsible for dimerization of the resolvase. The C-terminal 65 residues are responsible for DNA binding. Site-specific recombination involves the binding of resolvase to each *res* sequence, synpasis of the two *res* sequences, DNA cleavage, strand exchange and religation (Fig. 4.11). The *res* sequences must be in the appropriate head-to-tail orientation on the same DNA molecule, which has to be negatively supercoiled.

4.2.3 Inversion Systems

A sophisticated way of regulating gene expression involving site-specific recombination are the inversion systems, which have evolved as a mechanism to generate genetic diversity within a population. Inversion reactions have been termed programmed DNA rearrangements and usually occur at a low frequency. Inversion systems have been found on bacterial chromosomes, phages and plasmids. They consist of an invertible region flanked by inverted repeats and in general by one recombinase, sometimes called invertase. In several cases, site-specific inversion of DNA segments provides a simple on/off switch for genes located within or adjacent to the invertible region. In the simplest systems, the DNA rearrangements alternate a single segment of DNA, which generates one of two heritable traits. Examples discussed here are flagellar phase variation in *Salmonella*, host range variation in phages Mu and P1 and fimbriae variation in *E. coli*. More complex DNA rearrangements generate many different DNA isoforms, as examplified by plasmid *shufflons* (see next chapter).

Flagellar Phase Variation in *Salmonella*

The first example deals with flagella variation in some strains of *S. enterica* serovar *typhimurium* discovered in the 1940s. Cells can shift from making flagella composed of one flagellin protein, H1, to synthesize a different flagellin protein, H2.

Flagellin is the protein component of the flagellar filament. Flagellar variation is achieved by the reversible inversion of a 996-bp DNA segment which is bounded by two homologous chromosomal sites, *hixL* and *hixR* (Fig. 4.12). The *hix* sites are composed of two symmetrically related 12-bp halfsites separated by a 2-bp bridge. The invertible region contains a promoter that directs transcription of the gene encoding the H2 flagellin (*fljB*) and an adjacent repressor gene, *fljA*. In the H2-ON configuration, H2 flagella and the gene product of the *fljA* gene are produced, thereby repressing transcription of the *fljC* gene (which encodes the H1 flagellin). Upon inversion, the repressor gene is no longer transcribed and the H1 flagellin is expressed. Inversion of the DNA segment requires a site-specific recombinase (Hin, for *H in*version), which is encoded within the invertible segment and binds to and catalyzes site-specific recombination between the *hix* sites at either end of the segment. A third *cis*-acting DNA sequence was found to strongly influence the inversion rates: and this sequence is located within the coding region in the *hin* gene. Since this DNA sequence stimulates recombination regardless of its location and orientation within a plasmid substrate, it displays characteristics of a recombinational enhancer. The enhancer site contains two Fis binding sites, separated by 48 bp, that position the Fis dimers on opposite sides of the DNA helix. In addition, the histone-like protein HU is involved in inversion by facilitating DNA looping between *hixL* and the Fis binding sites and is required for high rates of recombination. The interaction of Hin, Fis and HU with their cognate binding sites forms a nucleoprotein complex termed the *invertasome*. This complex holds the invertible DNA segment in a topological configuration that allows Hin to catalyze the cleavage of four phosphodiester bonds. Rearrangement then occurs within the invertasome and Hin rejoins the exchanged DNA strands. Binding of the Fis protein increases the inversion rates about 1,000-fold, which occurs at a rate of around 10^{-4} per cell per generation. This low rate of inversion is believed to be primarily limited by the extremely low amounts of Hin protein present in cells.

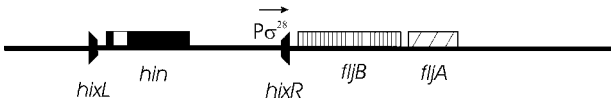


Fig. 4.12 The *Salmonella* flagellar phase variation system. The 996-bp H-inversion region of *S. enterica* serover *Typhimurium* controls flagellar phase variation. Inversion of the ó28-dependent promoter located between the two recombination sites *hixL* and *hixR* is catalyzed by the *hin* site-specific recombinase. In the present orientation, Pó28 transcribes the genes *fljB* and *fljA*, respectively coding for the H2 flagellin and a transcriptional repressor. The repressor protein prevents expression of the *fljA* gene encoding the H1 flagellin. Upon inversion of the H-region, *fljBA* are not expressed, resulting in the transcription of *fljA*.

Host Range Variation in Bacteriophages Mu and P1

Several phage DNAs have been found to contain DNA inversion systems similar to that of the *Salmonella* Hin system. The DNA of phage Mu contains a 3015-bp invertible region, designated the G segment, which is flanked by two 34-bp *gix*

sites which contain 12-bp sequences oriented as inverted repeats around a central region (Fig. 4.13A). The central, nonpalindromic 2-bp sequence confers directionality on the *gix* site and is bordered by the phosphodiester bonds that are broken and rejoined during recombination. Inversion of the G segment is catalyzed by the sequence-specific recombinase Gin (for *G in*version), which binds to the *gix* sequences. The recombination reaction is also stimulated by the Fis protein, which binds to a 60-bp enhancer sequence containing two 15-bp binding sites and located within the N-terminal region of the *gin* coding sequence. In a population of Mu-lysogenic cells, about 50% of the prophages carry the G segment in the (+) orientation, while the remaining cells carry it in the (–) orientation. What is the biological function of this inversion process? It has been shown that inversion affects the expression of two genes called *S* and *U*. While the promoter for both genes is located to the left of the G segment and the coding region of gene *S* starts outside of the invertible DNA segment, also to the left, the gene *U* is completely located within this segment (Fig. 4.13A). Depending on the orientation of the G segment relative to the flanking phage chromosome, two sets of genes are expressed: *S* and *U* in the (+) orientation and *S′* and *U′* in the (–) orientation. While the function of the *U* and *U′* genes remains elusive, genes *S* and *S′* code for the tail fiber. It follows that the tail fibers made when the G segment is in one orientation differ from those made when the segment is in the other orientation. Since tail fibers recognize receptors on the surface of sensitive bacteria, the two types of Mu phage particles should be able to adhere to different bacterial species. This is indeed the case; and Mu phages with S tail fibers attach to *E. coli* K-12 and *S. enterica* serovar *Arizonae* cells, while those containing S′ tail fibers adhere to other species such as *E. coli* C, *S. sonnei*, *C. freundii*, *E. carotovora* and *E. cloacae*. In summary, the biological function of the invertible G segment of phage Mu is to enlarge the host range. A similar invertible system is present in the DNA of phage P1, here called the C segment; and the inversion is controlled by the Cin (*C in*version) invertase (Fig. 4.13B). The 4.2-kb C segment is flanked by two large (0.62 kb) inverted repeats called *cixR* and *cixL* and this invertible DNA sequence

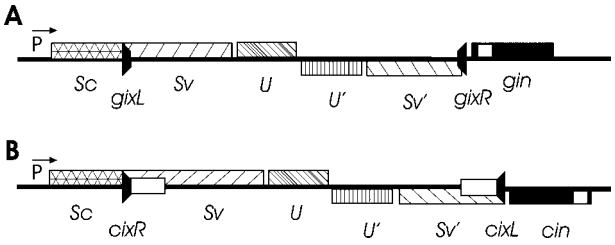


Fig. 4.13 The invertible DNA segments of bacteriophage Mu and P1. (A) G-inversion of phage Mu controls the alternating expression of tail fiber genes. The invertible segment is flanked by the inverted repeats *gixL* and *gixR*, and inversion is catalyzed by the Gin invertase. Depending on the orientation of the G segment relative to the flanking DNA, the genes *S* and *U* or *S′* and *U′* are expressed, where *S* and *S′* code for two different tail fibers. (B) C-inversion of phage P1 controls alternating expression of tail fiber genes, too. The genetic organization is similar to that of phage Mu.

is homologous to the G segment of phage Mu. Inversion controls the phage host specificity, with the C segment containing genes homologous in DNA sequence and function to S, U, S′ and U′ (genes *19* and *tfs*).

Fimbriae Variation in *E. coli*

The third example involves the expression of type I fimbriae (pili) encoded by *E. coli* K-12. As already mentioned, fimbriae play a key role in the attachment of Gram-negative bacteria to the host mucosa. In *E. coli*, type 1 fimbriae promote adherence of *E. coli* cells to mannose-containing receptors on host epithelial cells and have been implicated as an important virulence factor in urinary tract infections. The phase-variable expression of type 1 fimbriae arises from the reversible inversion of a 314-bp segment of chromosomal DNA that harbors the promoter for transcription of the *fimA* gene (Fig. 4.14). This gene codes for the type 1 fimbrial subunit; and inversion of its promoter sequence occurs via a site-specific recombination mechanism. Therefore, the orientation of the segment determines whether *fimA* is expressed or silent. The invertible element is known as the *fim* switch; and two recombinases of the integrase family, called FimB and FimE, are involved. These proteins have a number of interesting properties that make them stand out from the majority of integrase family members.

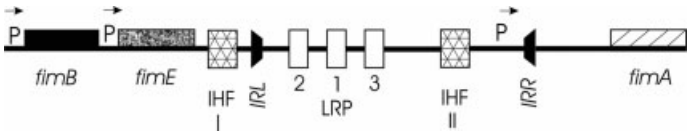


Fig. 4.14 Type I pili variation in *E. coli*. A 314-bp invertible region is flanked by the IRL and IRR sites and contains the promoter (P) responsible for the expression of the *fimA* gene encoding the typ I pilin. Two separate recombinase genes, *fimB* and *fimE*, are expressed from their own promoters. In addition, the location of binding sites for IHF and LRP are indicated that participate in the inversion reaction by forming the recombination complex.

FimB and FimE Act Independently of Each Other

FimB and FimE are the smallest members of the integrase family of recombinases. This makes them attractive subjects for analysis, because they may possess the minimum of structural information to function as recombinases.

Despite their small sizes, they have distinct biochemical activities at the *fim* switch. FimB inverts the switch in either the on → off or off → on direction with high efficiency. In contrast, the FimE recombinase has a strong preference for the on → off direction.

The FimB and FimE integrases have a strong dependency on IHF for normal function. There are two IHF binding sites at the switch and both are needed for normal inversion. They also require the *l*eucine-*r*esponsive regulatory *p*rotein, LRP, in order to invert the *fim* switch. There are three LRP binding sites within

the switch; and interactions between these and modulation of these interactions by L-leucine determine switching efficiency and link the operation of the switch to cellular physiology. The actions of both FimE and FimB are modulated in response to the growth temperature and the composition of the medium. FimB-promoted inversion acts optimally between 37 °C and 40 °C, while FimE-promoted inversion decreases as the temperature increases towards 37 °C. This infers a relationship between FimB and FimE which favors switching to the phase *on* state at 37 °C, a temperature appropriate for growth in humans and other mammals. In addition, FimE promotes the fimbriate to the afimbriate switch (on → off) at frequencies two orders of magnitude higher than FimB; and this may provide an efficient mechanism for ensuring rapid cessation of fimbrial synthesis in response to the appropriate environmental stimuli.

4.2.4
Shufflons

Shufflons are complex inversion systems (clustered inversion regions) which consist of multiple recombination sites and a site-specific recombinase, thereby allowing the expression of different proteins from the same gene. Shufflons are widespread among plasmids belonging to various incompatibility groups and all seem to be involved in the synthesis of pili. In addition, a shufflon has been identified in the phage P1-related plasmid p15B. Here, six *mix* DNA crossover sites are located within a 3.5-kb DNA inversion region (Fig. 4.15). Site-specific recombination between any two inverted *mix* sites catalyzed by the *min* invertase gives rise to six *S* gene products with different C-terminal segments. The *S* gene codes for a constant region of 594 amino acid residues, whereas the variable region fluctuates between 145 and 168 residues. The p15B *S* genes share sequence similarity with the tail fiber genes of various phages, although p15B is a defective phage. Therefore, the question arises: what could be the biological function of these different tail fiber-like proteins?

Fig. 4.15 The p15B Min inversion system. Any DNA sequence flanked by inverted *mix* sequences can invert giving rise to six different *S* gene products the functions of which remain elusive. T. Komano **1999**, *Annu. Rev. Genet.* **33**, 171–191; Fig. 4a.

4.2.5
Integrons

Integrons are natural genetic engineering systems that incorporate circularized open reading frames, called *gene cassettes*, and convert them to functional genes. The minimal integron system consists of an *integrase* gene (*intI*) of the tyrosine re-

combinase family and a *primary recombination* site, called *attI*, located proximal to the integrase gene. Integrons may contain one or more *gene cassettes* integrated at *attI* (Fig. 4.16). Gene cassettes contain a recombination site designated *attC* (or 59-base element) and a single ORF that encodes an adaptive function, mostly antibiotic resistance or other phenotypically detectable genes. The integrase catalyzes recombination between its cognate *attI* site and an *attC* site. The *attC* site exhibits a considerably length variance (from 57 bp to 141 bp) and is normally found associated with a single *orf* that encodes an adaptive function; and the *orf-attC* structure is termed a gene cassette. The *attI*-type sites and the 59-bp element-type sites have distinct architectures. The 59-bp element sites have diverse sequences and lengths but share regions of about 25 bp at each end that conform to consensus sequences. The consensus regions are imperfect inverted repeats of one another and each comprises a pair of inversely oriented integrase-binding domains, separated by a spacer of 7 bp or 8 bp. The arrangement of each consensus region is similar to that of the simple sites recognized by other integrases. In contrast, the sequences of the *attI* sites are not closely related to each other and do not share most of these features. The reactions catalyzed by the Int1 integrase encoded by class I integrons have been studied most extensively. IntI1 can catalyze both integrative and excisive recombination events and recombination between two 59-base elements, between *attI1* and a 59-base element and between two *attI1* sites has been documented. Recombination between the *attI* and *attC* sites leads to insertion of

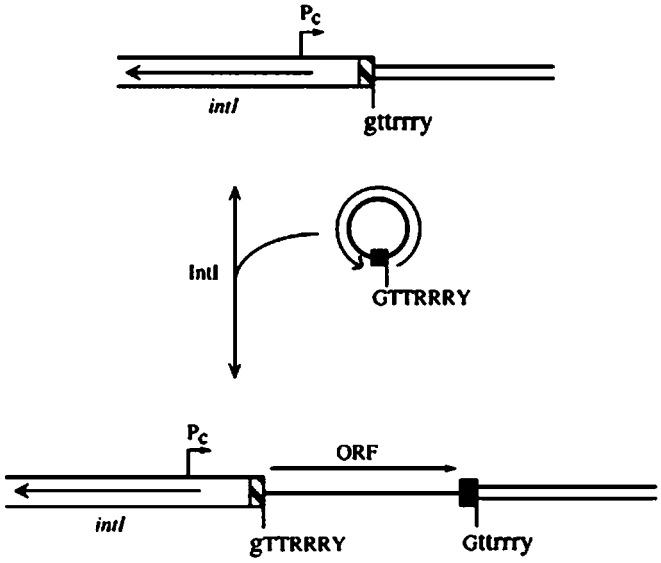


Fig. 4.16 Insertion of a circular gene cassette into an integron. *intI* codes for the integrase, the circular gene cassette normally for an antibiotic resistance gene, the promoter P_c is responsible for transcription of the gene in the cassette, and the hatched box symbolizes the 59-bp recombination site; the letters the core recombination sequence. R.M. Hall, et al. **1999**, *Ann. N.Y. Acad. Sci.* 870, 68–80; Fig. 1.

the gene cassette downstream of a resident promoter, P_c, within the integron that drives expression of the encoded product. The strongest promoter variant of P_c (TTGACA-N17-TAAACT) is six times more efficient than the derepressed P_{tac} promoter. Ordinary integrons are very compact: the largest known has eight cassettes, whereas most have between one and three cassettes.

The most notable gene cassettes identified within integrons are those conferring resistance to antibiotics. More than 70 different antibiotic resistance genes, covering most classes of antimicrobials presently in use, are structured as gene cassettes encoding resistance to aminoglycosides, β-lactams, chloramphenicol, trimethoprim, erythromycin, streptothricin, rifampicin and a variety of antiseptics and disinfectants. Significant reservoirs of multiresistant integron-containing strains have been found in animal populations, such as poultry and pigs. The stockpiling of these cassettes in integrons, to create *m*ulti*r*esistance *i*ntegrons (MRIs), has contributed substantially to the current dilemma in the treatment of infectious diseases, as integrons containing up to eight resistance cassettes have been found in clinical isolates with multiple resistance. Furthermore, integron systems are often found embedded within mobile DNA elements such as transposons and/or conjugative plasmids that can serve as vehicles for the intra- and interspecies transmission of the resistance genes that have been amassed by integrons. The proficiency of this partnership is confirmed by the marked differences in codon usage among cassettes within the same MRI, indicating that the antibiotic resistance determinants are of diverse origin.

Five classes of MRIs have been identified, based on the homology of the integrase genes. They share between 39% and 58% amino acid identity, which suggests that their evolutionary divergence has extended over a much longer period of time than the 50 years of the antibiotic era. The likely ancestors of MRIs are the recently discovered chromosomal *s*uper-*i*ntegrons (SIs) that have been identified in the genomes of diverse Gram-negative proteobacteria. The first SI was discovered in the *V. cholerae* genome. Clustered in one region of chromosome II, the *V. cholerae* SI spans 126 kb and harbors 214 open reading frames of mainly unassigned function in 179 cassettes. Its cassettes vary extensively in base composition and codon usage and appear to be of bacterial, viral and eukaryotic origin. The 179 cassettes of the *V. cholerae* array cannot, clearly, be expressed from one promoter, especially as they are not all oriented in the same direction. Moreover, the functions encoded by the SIs are apparently diverse and some are possibly related to pathogenesis. Other proposed functions, based on sequence similarity and experimental data, include the transport of small molecules, restriction modification, excreted lipase activity and plasmid-specific roles. Another example of a SI is the one found on the chromosome of *P. alcaligenes* called In55044. It is 18 kb in length, consists of 32 complete PAR elements (*P. alcaligenes r*epeat), 34 complete and two incomplete *orf*s.

MRIs and SIs share an identical structural organization, with an *intI* gene divergently transcribed from an upstream cluster of gene cassettes. Furthermore, the integrases of both MRIs and SIs are functional for cassette recombination and their integrases form a group related to but distinct from other members of the λ

family of site-specific tyrosine recombinases. Despite these commonalities, four major differences exist between MRIs and SIs:

1. MRIs typically contain fewer than six cassettes, with the largest described to date containing eight cassettes. SIs may contain hundreds of cassettes.
2. The cassettes found within MRIs typically code for antibiotic resistance genes whereas those of SIs are of mainly unknown function.
3. MRIs are commonly associated with mobile DNA elements. Conversely, the coevolution of SI integrases with their host genomes strongly suggests that SIs are sedentary.
4. Whereas *attC* sites of the gene cassettes of MRIs are highly variable in length and sequence, the *attC* sites of SI gene cassettes are closely related and species-specific. These relationships led to the proposal that MRIs evolved from SIs through the entrapment of *intI* genes and their cognate *attI* sites by mobile DNA elements. The subsequent harvesting of cassettes from various SI sources then led to the establishment of contemporary MRIs.

Any pathway suggested for gene cassette creation must account for all their features. One possible way involves reverse transcription of a mRNA molecule, as this would explain the fact that cassettes contain a complete gene with little flanking sequence and no promoter. The 59-base element could originate either from transcription terminators contained in the transcript or from some other structure that was added later. This hypothesis demands an reverse transcriptase; and such an enzyme has been found in several bacterial species in connection with retrons (see Section 2.2.1.2). Since no function has been ascribed to bacterial retrons, it is an interesting speculation that retrons might be involved in the formation of gene cassettes.

4.2.6 Homing Endonucleases

Homing is the transfer of a mobile genetic element to a cognate allele that lacks that element resulting in its duplication. Homing was first described for highly invasive introns (see Section 6.3.5) that carry a homing endonuclease; and later it was shown that inteins derived from precursor proteins by protein splicing (see Section 6.5.2) spread by the same mechanism. When an allele that harbors an intron or intein with a functional homing endonuclease coexists in the same cell as a gene that contains an allele without the mobile element, homing can be initiated by the endonuclease; and the intron/intein-free allele can be converted into an intein/intron-containing allele. By making a site-specific double-strand break in the intronless or inteinless alleles, these nucleases create recombinogenic ends which engage in a gene conversion process that duplicates the intron or intein (Fig. 4.17). During the repair of the double-strand break, the gene encoding the homing endonuclease and the surrounding sequences are copied into the cleavage site. This copying can be the result of homologous or illegitimate recombina-

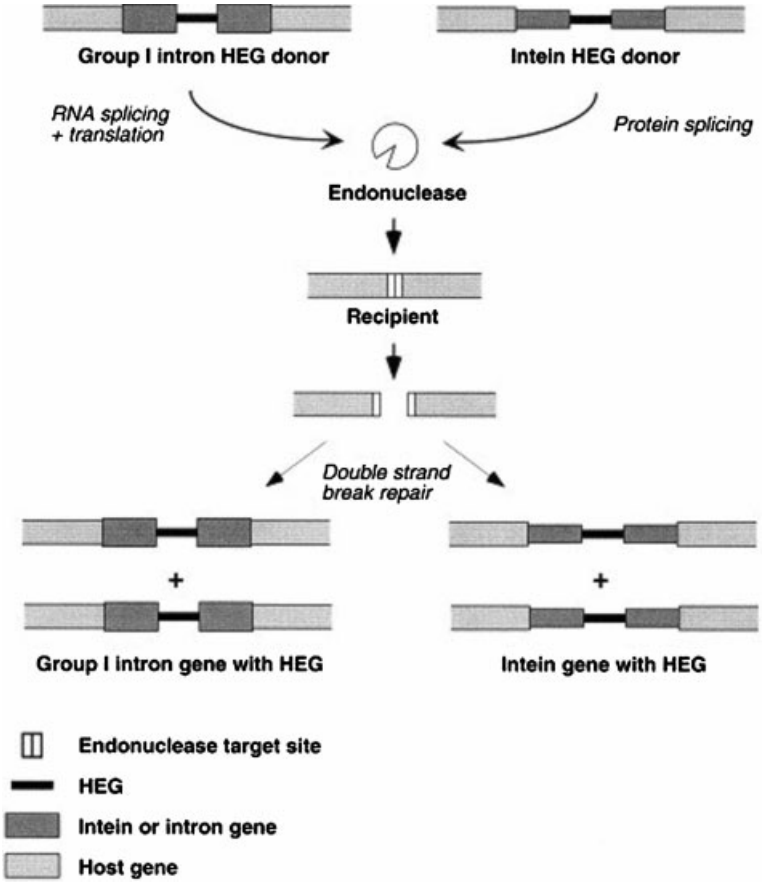


Fig. 4.17 Homing of a homing endonuclease gene (HEG). HEGs are located either in a group I intron or an intein. Group I introns are excised at the RNA level followed by their translation while inteins are spliced out from the precursor protein. Homing starts by a double-strand break in the recipient intronless or inteinless allele into which the HEG is inserted. F.S. Gimble **2000**, *FEMS Microbiol. Lett.* 185, 99–107; Fig. 1B.

tion. The former case results in homing, while the latter might result in a new mobile element if the endonuclease gene integrates into an existing intron or intein. Alternatively, the mobile element might be copied into a new target gene, and these elements are also able to transmit laterally within and between species.

While most homing endonucleases share with restriction enzymes the ability to make a site-specific double-strand break in the target DNA, they differ in structure, recognition properties and genomic location. Homing endonucleases:

- Fall within one of four families, characterized by the sequence motifs LAGLIDADG (this family is the largest, with more than 150 members reported to date), GIY-YIG, H-N-H and His-Cys box.

- Have recognition sequences that span 12–40 bp of DNA and usually do not require a complete match with the target sequence.
- Act as monomers or homodimers and, while some function independently of accessory molecules, others require associated proteins to regulate their activity and yet others form ribonucleoprotein complexes, wherein RNA molecules are integral components of the catalytic apparatus.
- Occur in all three biological kingdoms and are expressed in different compartments of the eukaryotic cell: nuclei, mitochondria and chloroplasts.

A few examples of homing endonucleases are given in Table 4.3.

Table 4.3 Homing endonucleases occur in all three biological kingdoms. Nomenclature: a three-letter genus/species designation consisting of the first letter of the genus and the first two letters of the species is followed by a Roman numeral to distinguish multiple enzymes from a single organism. Furthermore, intron endonucleases are characterized by the prefix I- (for intron), the intein endonucleases are characterized by the prefix PI- (for protein insert).

Species	*Name*	*Location*	*Gene/protein*
Lactococcus lactis	I-*Lla*I	Chromosome	LtrB
Mycobacterium tuberculosis	PI-*Mtu*I	Chromosome	RecA
Bacillus subtilis phage SP01	I-*Hmu*I	Phage	DNA polymerase
Pyrobaculum organotrophum	I-*Por*I	Chromosome	DNA polymerase
Thermococcus litoralis	I-*Tli*I	Chromosome	DNA polymerase
Aspergillus nidulans	I-*Ani*I	Mitochondrion	*cob*
Chlamydomonas eugametos	I-*Ceu*I	Chloroplast	Large rRNA
Physarum polycephalum	I-*Ppo*I	Nucleus	Large rRNA

4.3 Illegitimate Recombination

Illegitimate recombination is a class of recombination that takes place between sequences of little or no homology; and it results in DNA rearrangements such as deletions, translocations or insertions. Illegitimate recombination usually occurs at a low frequency, but is enhanced by X-ray, UV or other DNA-damaging agents and occurs preferentially between short regions of homologies on plasmid or chromosomal DNA. Illegitimate recombination can be classified into two classes, *s*hort-*h*omology-*i*ndependent *i*llegitimate *r*ecombination (SHIIR) and *s*hort-*h*omology-*d*ependent *i*llegitimate *r*ecombination (SHDIR). SHIIR occurs between sequences with virtually no homology and is mediated by DNA topoisomerases. SHDIR is induced by UV irradiation or other DNA-damaging agents and requires short regions of homology between recombination sites. These regions usually contain 4 bp and 10 bp of homologous DNA, respectively.

Transposable (mobile) elements are discrete segments of DNA capable of moving from one locus to another present either on the same or on a different replicon. They can be found in all three kingdoms of life and play fundamental roles as motors of genome plasticity. They induce various types of genomic arrangements and are the major cause of mutations. The smallest transposable elements consist of a single gene (often called *tnpA*) coding for *transposase*, an enzyme which catalyzes the transposition process, and *inverted repeats* (IRs) which flank the element and to which transposase binds. Transposition of a given element from one locus to another occurs by one of three mechanisms, called *conservative transposition*, *replicative transposition* and *formation of minicircles*. While in the first case, the element is excised from its resident locus and reintegrates into a new locus (also called the 'cut-and-paste' mechanism), the element becomes duplicated during the replicative transposition process. In both cases, integration at the new locus occurs with no sequence specificity by illegitimate recombination.

Transposable elements were discovered by Barbara McClintock in the 1940s when working with *Zea mays*; and she received the Nobel price in 1983 for her pioneering work. They were rediscovered independently in the late 1960s by two groups, Peter Starlinger and Heinz Saedler in Germany and Jim Shapiro in Cambridge (UK), both groups working with *E. coli*.

Based on their size and the genetic information they carry, bacterial transposable elements are classified into three groups, *i*nsertion (*IS*) *elements*, *transposons* and *transposable bacteriophages*. IS elements are small DNA sequences (600–1500 bp) that carry genetic information related to their transposition and its regulation. Transposons are larger (5–25 kb) and normally code for one or more antibiotic resistance marker. Transposable phages are even larger (about 40 kb) and use the mechanism of transposition for propagation of their DNA. *E. coli* phage Mu is the paradigm for this group of mobile elements.

Transposition events are rare, occurring at a frequency of one in 10^{-5} to one in 10^{-9} cells per generation. This low transposition frequency makes good biological sense, since a high frequency would raise the possibility of a lethal insertion mutation, resulting in the loss of the mobile element. Several mechanisms have been described leading to the low rate of transposition, including a tight regulation of expression of the *tnpA* gene (e.g., by the antisense strategy in Tn*10* or by a transcriptional repressor in Tn*3*), low activity of the transposase (Tn*5*) and complexing with an inhibitor protein (Tn*5*). All these strategies have evolved to ensure a balance between transposition and host survival.

4.3.1 Insertion Sequence Elements

Insertion sequences (IS) are small, generally cryptic transposable elements, 0.7–2.5 kb in length, which are ubiquitous in bacterial genomes and contribute significantly to spontaneous mutagenesis in bacteria. Insertion of an IS element often leads to gene inactivation and strong polar effects on expression of the downstream gene(s) belonging to the same trancriptional unit. However, certain IS ele-

ments carry promoters or other protein-binding sequences, such that an IS insertion can activate cryptic genes (see Section 6.1.5) or alter the expression of adjacent genes. Activation of neighboring genes can also be due to the formation of hybrid promoters when insertion of an IS results in placing an outwardly directed –35 promoter box located in the terminal inverted repeat at the correct distance from a resident –10 box. Furthermore, IS elements act as substrates for recombination pathways leading to chromosomal rearrangements, such as inversions and deletions. More than 700 IS elements have been identified from 171 bacterial species and classified into about 20 families, based on homology among their transposases (see http://www-is.biotoul.fr). IS elements in most of these families have terminal inverted repeat sequences, 10–40 bp in length, which can often be divided into two functional domains. Domain I includes the two or three terminal base pairs and is involved in the cleavages and strand transfer reactions leading to transposition of the element. Domain II is positioned within the IR and is involved in transposase binding. Another general feature is that, on insertion, most IS elements generate short directly repeated sequences of the target DNA flanking the IS. Attack of each DNA strand at the target site by one of the two transposon ends in a staggered way (see below) during insertion provides an explanation for this observation. The length of the direct repeat (between 2 bp and 14 bp) is characteristic for a given IS element. Most of the IS elements belonging to the IS*110*/IS*492* family have no terminal IR and do not generate duplication of the target site sequence upon transposition.

Most IS elements contain only one open reading frame coding for the transposase (the *tnpA* gene); and a few carry additional coding information. As already mentioned, transposition is a rare event and depends solely on active transposase. Transposase promoters are generally weak and many are partially located in the terminal inverted repeats, enabling their autoregulation by the transposase. Additional regulation can be exerted by a transcriptional repressor (IS*1*, IS*2*), translational inhibitors (antisense RNA in the case of IS*10*), impinging transcription (IS*10*, IS*50*), programmed translational frameshifting (IS*1*, members of the IS*3* family; see Section 6.4.7.2) and transposase stability. While the intrinsic promoters are weak, the transposase genes are at risk to become activated from a strong promoter located outside of the element. Many IS elements have evolved mechanisms that attenuate activation of their *tnpA* genes by impinging transcription from an outside promoter. One such mechanism consists of sequestering translation initiation signals (Shine–Dalgarno sequence and/or start codon) in an RNA secondary structure. Another target is the transposase protein. In the case of IS*903*, the halflife of the transposase is controlled by the Lon protease (see Section 7.5.1).

4.3.2 Transposons

Bacterial transposons have a length from 5–25 kb and code for one or more phenotypic genes. They are classified into two groups based on their structure, *simple* transposons and *composite* transposons. Simple transposons encode transposition

functions and carry inverted repeats at their ends, which are the recombination targets. The determinants encoded by these elements include antibiotic resistance genes, virulence factors and catabolic genes. These elements are structurally diverse and a well studied group of elements is the Tn*3* family. Composite transposons contain two IS elements bounding, as either direct or inverted repeats, a DNA segment which encodes auxiliary determinants such as antibiotic resistances and virulence factors. The flanking IS components provide the recombination machinery, i.e. transposase and *cis*-acting sites at the termini of the ISs, which mediate translocation where the outside ends of the IS segments collaborate to translocate the entire composite element.

4.3.3
Transposing Bacteriophages

The temperate *E. coli* phage Mu is the paradigm of transposing bacteriophages. Its name derives from its ability to integrate into numerous sites in the *E. coli* chromosome and cause mutations by insertional inactivation (Mu is an abbreviation of *mu*tator). Two pathways for Mu transposition are generally distinguished. Upon infection of a sensitive host cell, the linear double-stranded DNA is converted into a noncovalently closed circular form with the help of an injected protein which also protects the ends from exonucleases. Next, the DNA integrates at random in the host chromosome, and this process is called *conservative integration*. After about 10 min, each infected cell has to decide whether the Mu DNA goes into the lysogenic or lytic cycle. This is dictated by the amount of Mu repressor encoded by the *c* gene present. If its amount is high, the lysogenic cycle results and the integrated Mu DNA molecule is passively replicated as a *prophage* where the Mu repressor prevents expression of all the genes needed for the lytic development. If the amount of repressor stays low, the lytic cycle is initiated. The original Mu DNA molecule is replicated by *in situ* replication and the copy inserts at another site (*replicative transposition*). In the next round of replication, both copies serve as templates for *in situ* replication; and the two new copies integrate somewhere else. The replication cycles continue until about 100 copies of Mu DNA are scattered around the host chromosome. While replication goes on, transcription of the late genes coding for the coat proteins occurs in parallel. After about 50 min, replication ceases and the DNA molecules become packaged into Mu proheads. Packinging enzymes recognize a *pac* site located very close to the left *attL* end and they catalyze a double-strand cut 50–100 bp outside of *attL* within the host DNA. Using this free *attL* end, packing into an empty prohead starts and follows the full head packaging mechanism. This means that the head becomes completely filled with DNA. Since the length of the Mu DNA (37 kb) is not sufficient to fill the head, a second cut occurs 1.0–1.5 kb to the right of *attR*, at the right end of the Mu DNA. It follows that each Mu DNA molecule is flanked by foreign DNA, 50–100 bp on the left side and 1.0–1.5 kb to the right side. Since each DNA molecule is derived from a different location on the host chromosome, it follows that the outside ends are also different.

4.3.4 Mechanisms of Transposition

As mentioned above, mobile elements move by one of three different mechanisms, depending on the element.

Conservative (Nonreplicative) Transposition: the Cut-and-paste Mechanism

The cut-and-paste mechanism seems to be the most prominent one and appears to be very popular in eukaryotic organisms as well. The biochemistry of reactions has been examined in detail for Tn*5*, Tn*7*, Tn*10* and IS*911*. The mobile element is excised completely from its location by the action of the cognate transposase, which makes double-strand breaks at each transposon end (Fig. 4.18). The concerted action of the released 3′ transposon ends on the two strands of a duplex DNA target generates a simple insertion of the excised, unreplicated element. The two short single-strand gaps are repaired.

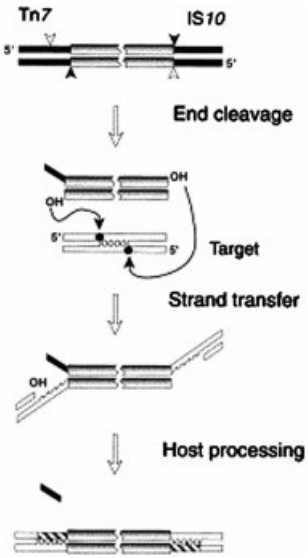


Fig. 4.18 Biochemical steps leading to nonreplicative transposition. The shaded rectangles represent the two ends of the DNA strands of the mobile element, black and white rectangles the flanking donor and target DNA sequences, respectively. Cleavage sites for Tn*7* and Tn*10* are indicated by open and closed short arrows, respectively. Curved arrows indicate the nucleophilic attack transferring the 3′ OH ends on staggered phosphates of the target DNA symbolized by black dots, while the zig-zag lines represent the few target nucleotides that are duplicated during the transposition process, the cross-hatching replication events that complete transposition after complex dissociation. B. Hallet, D.J. Sherratt **1997**, *FEMS Microbiol. Rev.* 21, 157–178; Fig. 4.

Replicative Transposition

Mobile elements such as bacteriophage Mu and the transposon Tn*3* make use of the replicative transposition mechanism (Fig. 4.19). The elements are nicked only once at each end and remain connected at the 5′ ends to the flanking DNA sequences. In the second step, the 3′ OH ends attack the target site and join both ends of the element to staggered phosphates of the two target DNA strands. Complete replication of the mobile element leads to the formation of a cointegrate, the final product of replicative transposition. This cointegrate may be resolved by site-

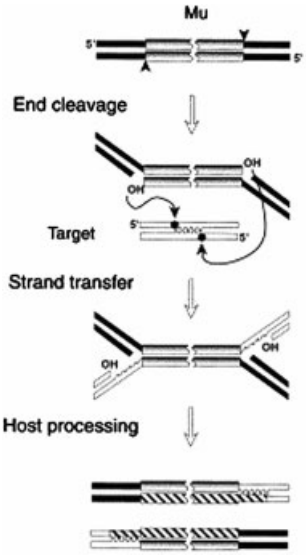


Fig. 4.19 Biochemical steps leading to replicative transposition. For explanation of the different symbols see legend to Fig. 4.18. B. Hallet, D.J. Sherratt **1997**, *FEMS Microbiol. Rev.* 21, 157–178; Fig. 4.

specific recombination between two DNA sequences called *res* sites to restore the initial donor molecule and yield a target molecule with the mobile element.

Formation of Minicircles

IS elements IS*2,* IS*3* and IS*911* transpose by the formation of minicircles. Transposition is initiated asymmetrically by single-strand cleavage at just one IS 3′ end. In the case of IS*2,* terminal cleavage initiates exclusively at the right inverted repeat (IRR), called the donor end. This free 3′ end then attacks a unique target adjacent to the other IS end (IRL), the target end, joining to the same DNA strand and thereby circularizing one strand of the IS. This converts the donor replicon into a figure-of-eight intermediate with the fourway junction held together by the single strand end-to-end connection (Fig. 4.20). Replication and/or repair of the figure-

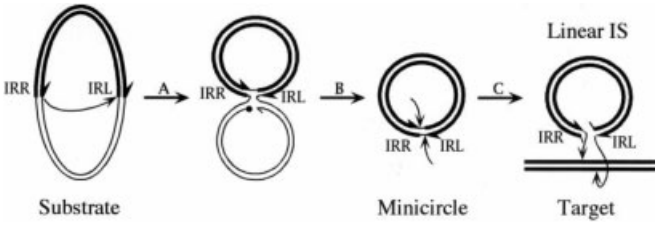


Fig. 4.20 Pathway of IS*2* transposition. (A) Formation of the figure-8 intermediate by asymmetric cleavage of IRR and its strand transfer to IRL target (B) Resolution of the figure-8 by replication or repair to produce the IS*2* minicircle (C) Cleavage of the minicircle junction to form the linear IS which can insert into a DNA target. L.A. Lewis **2001**, *Mol. Microbiol.* 42, 887.

of-eight generates a double-stranded IS minicircle with the two abutted ends separated by a short spacer of 1–2 bp. Next, the minicircle junction with its head-to-head IS ends serves as the substrate for transposase-mediated cleavages at both 3′ ends.

4.3.5 Conjugative and Integrative Elements

Conjugative and integrative elements are defined as DNA sequences normally present on the bacterial chromosome being able to excise and catalyze their transfer from the donor to a recipient cell. These elements can be divided into two groups, *conjugative transposons* and *integrating conjugative elements*. While conjugative transposons integrate with little specificity and are able to transpose within a cell, integrative conjugative elements recognize one specific integration site and therefore never move within the cell already harboring a copy. Tn*916* from *E. faecalis* serves as a paradigm for conjugative transposons, while pSAM2 from *S. ambofaciens* is a representative of the integrating conjugative elements.

*C*onjugative *tran*sposons (CTns) are genetic elements that encode their own integration, excision and transfer functions. They are remarkably promiscuous and are capable of being transferred across large phylogenetic distances. They are important clinically, as they are one of the major vectors involved in the spread of antibiotic resistance among bacterial pathogens. CTns can be defined as integrated self-transmissible elements that excise from the chromosome to form a circular intermediate structure before transferring themselves to a recipient cell by conjugation. In the recipient, the transferred circular form integrates into the recipient's chromosome. CTns have been found in a variety of bacteria, but have been most extensively studied in the Gram-positive cocci and *Bacteroides* subspecies. The best studied *Bacteroides* conjugative transposons are CTnERL and CTnDOT. The basic structures of these two CTns are the same, but CTnDOT carries an additional 13-kbp region that contains an antibiotic resistance gene, *ermF*. Integration of CTnDOT into the *Bacteroides* chromosome requires an integrase (IntDOT) and a segment of DNA that contains the CTnDOT joined ends (*attDOT*). Integrase gene expression is constitutive. In the case of phage λ and the conjugative transposon Tn*916*, excision of the element from the chromosome requires both an integrase and an excisionase (Xis). The Xis proteins of phage λ and the conjugative transposon Tn*916* are both small basic proteins that have the same function in excision, but their amino acid sequences have no significant similarity to each other.

The Conjugative Transposon Tn*916*

Tn*916* and the other conjugative transposons do not encode pili; and the mechanism of cell-to-cell contact remains unknown. The first step in Tn*916* conjugative transfer is an excision event catalyzed by the transposon encoded integrase and excisionase leading to a covalently closed circular molecule with an attachment site

attTn resulting from recombination between the two attachment sites *attL* and *attR* which flank Tn*916*. The second step encompasses the transfer of the excised transposon starting with a nick at the *oriT*, transfer of single-stranded DNA to the recipient cell and completion to double-stranded molecules in both the donor and recipient cell. The last step is the integration of the two transposons by the tyrosine-type integrase between *attTn* of the circular form and the bacterial chromosome, preferentially into A+T-rich sequences.

The Integrating Conjugative Element pSAM2

The pSAM2 element is integrated into a unique site of the conjugative chromosome, the 3′ end of the gene encoding $tRNA^{Pro}$. A tricistronic operon of this element codes for the genes *repSA* (rolling-circle replicase), *xis* (excisionase) and *int* (tyrosine recombinase). Expression of this operon is under the positive control of the *pra* gene, whose transcription is repressed in the absence of recipient cells. In the present case, recipient cells are vegetative mycelia, septated, multinucleoid, branched hyphae that grow into the solid nutrient substrate. After physical contact between donor and recipient mycelia, pSAM2 becomes excised, the excised copy undergoes rolling-circle replication and transfer occurs as double-stranded DNA catalyzed by *traSA*. The TraSA protein is related to SpoIIIE from *B. subtilis* acting as a double-stranded DNA pump during the initiation of sporulation. The transferred copies of pSAM2 are then replicated within the recipient mycelium, followed by intramycelial transfer and integration at the $tRNA^{Pro}$ gene.

Most of the integrating conjugative elements insert unique sites, usually the 3′ end of genes coding for tRNAs. This behavior prevents transposition within the chromosome when a copy is already present.

4.3.6 Nonhomologous Endjoining (NHEJ)

One major mechanism of double-strand repair in eukaryotes is nonhomologous endjoining (NHEJ), in which the two ends of a broken chromosome are rejoined directly using only the base pairing information of the two ends themselves. Later, this mechanism has also been found in bacteria through alignment of the protein sequences involved in NHEJ first. The enzymes have lately been purified from *M. tuberculosis* and shown to work as expected. NHEJ starts with the binding of a protein of the YkoV family (name derived from the *B. subtilis* gene) to both ends of the broken chromosome and synpasis (Fig. 4.21). Next, a helicase (RecQ in *E. coli*) and a nuclease (YkoU in *B. subtilis*) catalyze the terminal processing followed by ligation of the nicks through a ligase. Since, for instance *B. subtilis* codes for this enzyme complement, one might ask whether cells are capable of using NHEJ to circularize transformed linear plasmids. NHEJ might be especially efficient in stationary-phase cells to counteract double-strand breaks induced by stress factors such as heat and desiccation and might also participate in stationary-phase mutagenesis (see Section 5.1.5).

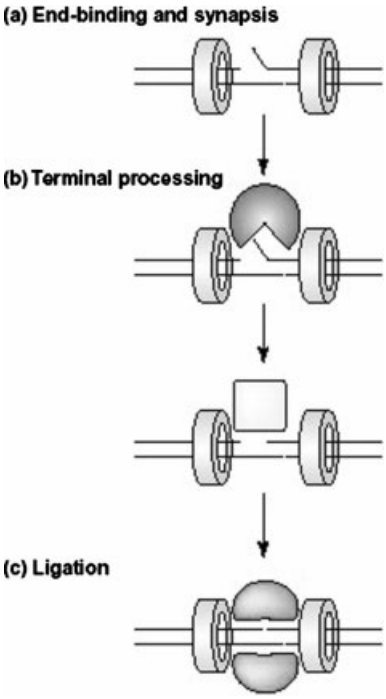


Fig. 4.21 Nonhomologous endjoining of broken DNA molecules. The ends of a broken DNA moelcule are recognized by a protein (YkoV in *B. subtilis*) and brought into vicinity. Other proteins are involved in terminal processing (helicase, exonucleases), fill-in (DNA polymerase) and ligation. T.E. Wilson, et al. **2003**, *Trends Biochem. Sci.* 28, 62–66; Fig. 1.

Suggested Reading

M. Belfort, R.J. Roberts **1997**, Homing endonucleases: keeping the house in order, *Nucleic Acids Res.* 25, 3379–3388.

V. Burrus, et al. **2002**, Conjugative transposons: the tip of the iceberg, *Mol. Microbiol.* 46, 601–610.

M.M. Cox **2003**, The bacterial RecA protein as a motor protein, *Annu. Rev. Microbiol.* 57, 551–577.

F.S. Gimble **2000**, Invasion of a multitude of genetic niches by mobile endonuclease genes, *FEMS Microbiol. Lett.* 185, 99–107.

R.M. Hall, et al. **1999**, Mobile gene cassettes and integrons in evolution, *Ann. N.Y. Acad. Sci.* 870, 68–80.

L. Haren, et al. **1999**, Integrating DNA: transposases and retroviral integrases, *Annu. Rev. Microbiol.* 53, 245–281.

I.R. Henderson, et al. **1999**, Molecular switches – the ON and OFF of bacterial phase variation, *Mol. Microbiol.* 33, 919–932.

T. Komano **1999**, Shufflons: multiple inversion systems, *Annu. Rev. Genet.* 33, 171–191.

D.A. Rowe-Magnus, D. Mazel **2001**, Integrons: natural tools for bacterial genome evolution, *Curr. Opin. Microbiol.* 4, 565–569.

D. Summers **1998**, Timing, self-control and a sense of direction are the secrets of multicopy plasmid stability, *Mol. Microbiol.* 29, 1137–1145.

5
Origin of Mutations and Repair of DNA Lesions

A mutation is defined as any heritable change in the DNA sequence of an organism. Mutations are a characteristic of all living systems and provide the material for natural selection. Mutations occur spontaneously as a result of replication errors and interactions between the cellular metabolism and the DNA, defining the *spontaneous mutation rate* which cannot be reduced. Typically, spontaneous mutations occur at very low rates: about 10^{-10} mutations per base pair per cell division in both pro- and eukaryotic organisms. This spontaneous mutation rate can be increased by treatment of the cells with *mutagens*, leading to *induced mutations*. UV irradiation and certain chemicals act as potential mutagens. Most mutations are *point mutations*, changing one base pair to another. Addition of one or two or deletion of one or two base pairs leads to *frameshift mutations*. Rearrangements within the genome include deletions, duplications, insertions and inversions. To achieve a low mutation rate, cells must use numerous mutation-avoidance strategies. These strategies include a high accuracy of DNA replication and the maintenance of the DNA in an error-free state by eliminating DNA-damaging agents or repairing DNA damage once it has occurred. DNA repair systems include the *methyl-mediated mismatch* repair system, the *nucleotide excision* repair system, pathways of *base excision* repair and, last but not least, the *SOS response*.

5.1
Classes of Mutations

Three different classes of mutations are distinguished:

1. point mutations, including small deletions and insertions
2. frameshift mutations
3. rearrangements, including insertions, duplications, deletions and inversions.

Point Mutations

Most point mutations involve single base pairs; and here two subclasses are distinguished. If one purine is replaced by another purine, or one pyrimidine by the other, this alteration is called a *transition*. If a purine is replaced by a pyrimidine

Dynamics of the Bacterial Chromosome. Structure and Function. W. Schumann
Copyright © 2006 Wiley-VCH Verlag GmbH & Co. KGaA
ISBN 3-527-30496-7

and vice versa, this is termed a *transversion*. As to the consequences of a point mutation, there are four different outcomes. If the mutation leads to the replacement of a totally different amino acid (e.g., alanine by proline), this will lead to a *missence mutation* and the protein affected by this replacement might lose part of its activity. Replacement of a sence by a nonsense codon will lead to a *nonsense mutation*. Depending on the position of the nonsence mutation within the coding region, this will lead to the formation of a truncated protein which will be detected by ATP-dependent proteases such as Lon and degraded. If an amino acid of one group is replaced by another from the same group (e.g., leucine by isoleucine), this is termed a *conservative mutation* and will not affect the structure and activity of the protein unless this amino acid residue is part of the active site. Mutations can also occur within replacement of one by another amino acid. This behavior is based on the fact that most amino acids are encoded by more than one codon, very often exhibiting variations in the third position. One example is the amino acid proline, encoded by the four codons CCU, CCC, CCA and CCG. If the wild-type codon is CCU and the U is mutated either to C, A or G, a proline will be inserted during translation. This last type of mutation has been designated a *silent mutation*. Mutations can also mutate back to restore the original codon and this event is called *reversion*. But the activity of a mutated protein can also be restored by a suppressor. Here, *intragenic* and *extragenic* suppressors are distinguished. Intragenic suppressors result from a second mutation within the mutated gene, restoring the activity of the protein encoded by that gene. In contrast, extragenic suppressors occur in a different gene, restoring the activity of the first. One example is given by two different proteins A and B forming a complex. A point mutation in gene *A* leads to the synthesis of a protein A′ unable to interact with protein B. A point mutation in gene *B* results in the synthesis of a protein B′ now able to form a complex with its partner protein A′. The point mutation within gene *B* is the extragenic suppressor. Isolation of extragenic suppressors play an important role in identifying proteins which interact with each other.

Frameshift Mutations

Frameshift mutations occur as a result of the insertion or deletion of one or two base pairs which will change the reading frame into another. If translation continues in the new reading frame, different amino acid residues will be added; and in most cases this new reading frame ends after a few codons in a nonsense codon, leading to release from the ribosomes. Again, the location of the frameshift dictates the fate of the polypeptide chain. If it is close to the 5′ end of the gene, the polypeptide will have no function and is degraded. If it occurs close to the 3′ end, it may retain part or its complete function. There are a few cases, where frameshifting is genetically regulated to produce either two different proteins or one with and the other without a function (see Section 6.4.7.2). There are two major mechanisms responsible for the generation of frameshift mutations: slippage during replication and intercalation of acridine dyes. Replication slippage involves homopolymeric tracts, for instance several consecutive G residues (Fig. 5.1, upper

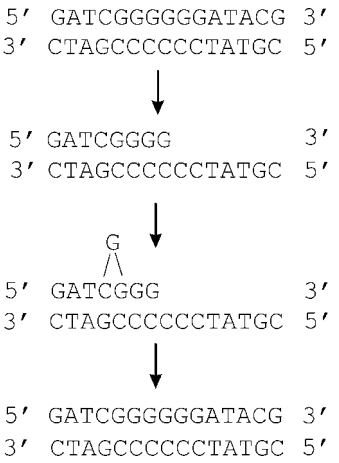


Fig. 5.1 Mechanisms responsible for the generation of frameshift mutations. (Upper part) Replication slippage. The replison misaligns at a run of identical nucleotides. (Lower part) Intercalating acridine dyes. Acridine dyes can intercalate into a DNA strand causing slippage during replication.

part). Here, the replisome misaligns, which can lead either to the deletion of one basepair (–1 frameshift) or to the insertion of one basepair (+1 frameshift). Molecules of the acridine family are usually planar molecules which are able to intercalate between bases in the same strand of the DNA, thereby increasing the distance between the bases and leading to misalignment, too. Intercalation of a dye molecule will force two of the bases apart, causing slippage of the two strands with respect to each other. Whether a deletion or addition of a base occurs is dependent on which strand slips. If the dye is intercalated in the newly synthesized strand, this strand might slip backward, leading to the deletion of one basepair during the next round of replication (Fig. 5.1, lower part). If the dye inserts into the template strand prior to replication, the newly synthesized strand might slip and incorporate an extra nucleotide. The frameshift mutagens include acridine dyes, ethidium bromide and aflatoxins produced by fungi.

Rearrangements

As mentioned above, there are different types of rearrangements. *Insertions* occur in most cases by a mobile element which translocates from one site to another on the same chromosome or from another replicon such as a plasmid or a bacteriophage into the chromosome. Since most mobile elements insert with low (if any) sequence specificity, they might insert into genes and lead to an insertion mutation. If this gene is essential, the insertion will lead to the death of the affected bacterial cell. *Deletions* leading to the loss of DNA sequences can be grossly divided into small and large deletions. Small deletions involving a few nucleotide pairs might occur during replication by slipped mispairing in tracts of any one of the four nucleotides. Large deletions are the products of a homologous recombination event between two identical or almost identical DNA sequences, e.g., two IS elements present in the same orientation on the chromosome. During the recombination event, one copy and the DNA sequences located between the two ele-

ments are deleted and are then lost from the cell. While most deletions occur spontaneously, there are a few reported cases where the deletion event is genetically regulated (see Section 2.4). *Inversions* involve identical or almost identical sequences occurring in an indirect orientation towards each other. During homologous recombination involving both sequences, the DNA located between them is inverted. As for deletions, there are many cases of genetically regulated inversion events described in detail under Section 4.2.3.

5.2 Origin of Mutations

5.2.1 Spontaneous Mutations

Analyses of spontaneous mutations along the chromosome have revealed that they do not arise randomly in the genome sequence but rather possess a strong bias in their site of distribution. Each type of mutation shows a pattern of *hotspot* and *cold spot* sites in a given sequence, suggesting that spontaneous mutagenesis depends on the context of the DNA sequence. It can be assumed that most of the events in spontaneous mutagenesis are affected by DNA topology and DNA transactions, such as transcription, replication and recombination. The molecular mechanisms of spontaneous mutagenesis have been analyzed by the study of *mutator* and *antimutator* genes in *E. coli* and by biochemical studies of their gene products. Mutator genes increase the spontaneous mutation rate, while antimutators decrease them.

Mutagenesis is always a two-step process. During the first step, the *premutagenic step*, some premutagenic damage occurs in the DNA leading to a premutation. If this damage is not repaired before the replication fork arrives, it will be converted into a true mutation during the second step, the *mutagenic step*. The major cause of premutagenic damage are:

- replication errors made during normal DNA synthesis
- spontaneous DNA lesions and mutagenic substrate nucleotides
- replication errors induced by mutagenic nucleotides
- replication errors made during normal DNA synthesis.

Three types of replication errors have been well characterized.

Single-base Mispairs Leading to a Base Substitution

These replication errors consist of only undamaged nucleotides and are fixed into a mutation at the next round of replication. At the DNA replication step, several levels of fidelity have been found. In *E. coli*, DNA polymerase III holoenzyme has an error rate in the order of 10^{-6} to 10^{-7} per base pair. This fidelity results from two separate steps: base selection at the insertion step and editing (proofreading)

by the 3′→5′ exonuclease associated with the DNA polymerase. Following replication, the DNA is scrutinized by a methyl-dependent mismatch-correction system (see below).

Single-base Bulges Leading to a Single-base Frameshift

Misalignment of the growing chain with the template sequence may occur via simple backwards slippage of the terminus, thereby promoting the formation of single-base bulges and resulting in a +1 frameshift. Less frequent are insertions or deletions of a single or multiple unit of dinucleotide repeats, which may involve a simple slippage event during DNA synthesis.

Multiple-base Mismatches Leading to a Sequence Substitution

Template switching potentially generates a multiple-base mismatch. This third category of replication error has been proposed as a mutagenic process leading to sequence substitution, ranging in length from 2 nucleotides to about 20 nucleotides. Based on the definitive requirement of pseudo-inverted repeats for sequence substitution mutagenesis, intra- or interstrand template switching between repeats was proposed for mutagenesis. However, the molecular mechanisms of template switching are poorly understood.

Based on the rate of uncorrected replication errors, four to six mutations are expected to be produced per chromosome per DNA replication cycle.

Spontaneous DNA Lesions and Mutagenic Substrate Nucleotides

Less is known about the DNA lesions that potentially induce spontaneous mutations. Active oxygen species are produced during aerobic growth and attack DNA, to produce a wide variety of lesions leading to an estimated 3000–5000 oxidative DNA lesions/cell/generation. Free nucleotides are attacked more efficiently than DNA, and both 8-oxo-dGTP and 2-OH-dATP possess extraordinarily strong mutagenicity and are assumed to be the most powerful source of spontaneous mutations. Methylation and hydrolytic decomposition of DNA, such as depurination and cytosine deamination, cause a variety of endogenous DNA lesions.

Replication Errors Induced by Mutagenic Nucleotides

There are two additional pathways in which replication errors are induced during DNA synthesis. One is the misinsertion of a mutagenic nucleotide. When 8-oxo-dGTP is incorporated opposite either C or A residues, a mispair results which can be converted to a premutation in the next round of replication. Fixation of this premutation into a base-substitution mutation requires a further round of DNA replication. The other pathway is the misinsertion of a normal nucleotide opposite a miscoding type of spontaneous DNA lesions. One example is the misincorporation of a dATP opposite an 8-oxoG residue. The resulting A:8-oxoG mispair is a

premutation leading to a G:C→T:A transversion at the next round of DNA replication.

5.2.2 Induced Mutations

That the spontaneous mutation rate can be enhanced experimentally was shown by H. J. Muller in 1927 for the first time. He treated *D. melanogaster* with X-rays and UV light and observed an increase in phenotypes. Later, it was shown that many chemicals of various types increase the frequency of mutations. Broadly, these mutagenic agents are called *mutagens*. How do they act?

5.2.2.1 Ionizing Radiation Mutagens

A number of different types of high-energy radiation produce ionizations resulting in the generation of charged atomic and molecular particles by displacement of electrons. This property is shared by short-wavelength electromagnetic radiation (including X-rays and the still shorter-wavelength γ-rays) and by beams of high-velocity atomic particles, such as neutrons and helium nuclei (α particles). All ionizing radiation induces chromosome breaks, leading to DNA rearrangements and point mutations. The primary effect of UV light consists to a very large extent of the formation of crosslinks between pairs of pyrimidines, called pyrimidine dimers. Thymine dimers are the most stable and plentiful type in UV-irradiated DNA, but cytosine dimers and mixed thymine-cytosine dimers also occur with frequencies depending on the DNA base composition. Pyrimidine dimers can be resolved by the enzyme photolyase in the presence of visible light. The repair mechanism is designated *photoreactivation* (see below).

5.2.2.2 Chemical Mutagens

The first chemical to be shown to induce mutations was the bifunctional alkylating agent bis(2-chloro-ethyl) sulfide by C. Auerbach in 1946. Since this initial discovery, the list of chemical mutagens has grown enormously and a wide range of compounds has been shown experimentally to be mutagenic to different degrees. In order to be a mutagen, a chemical has to fulfill one or other of the following three criteria:

- it must react chemically with DNA; or
- it must be convertible metabolically into a compound which will react with DNA; or
- it must mimic in some way a normal DNA base or basepair so as to interfere with replication.

Alkylating Agents Chemicals that will alkylate DNA bases include some of the most powerful mutagens known. Two different types of alkylating agents are distinguished. *Monofunctional* alkylating agents have just one alkyl group to donate,

usually methyl or ethyl. Examples are *e*thyl*m*ethane sulfonate (EMS) and *N*-methyl-*N'*-nitro-*N*-nitrosoguanidine (NG). *Bifunctional* agents contain two reactive groups per molecule and hence can simultaneously link to two DNA bases and form crosslinks between different DNA chains or different parts of the same chain. Mustard gas is one example of a bifunctional chemical. The main targets for alkylation in the DNA are the nitrogen atoms of the bases: the N-7 position of guanine, the N-1 and N-3 positions of adenine and the N-1 position of pyrimidines. In addition, the O-6 atom of guanine is also alkylated to a significant extent.

Polyaromatic Compounds A potent mutagenic agent is the polyaromatic compound benzopyrene and other compounds of related structure. These compounds are rather inert chemically and will not themselves react with DNA bases. But they can be transformed by oxidative enzymes of the cellular respiratory system to epoxide derivatives. The epoxide ring is very reactive and will attack vulnerable atoms in the DNA duplex, probably the same ones that are subject to alkylation.

Hydroxylamine and Nitrous Acid Hydroxylamine specifically attacks only cytosine among the DNA bases, creating a cytosine derivative with the hydrogen-bonding properties of thymine, thus causing G-C→A-T transitions. Hydroxylamine is not an effective mutagen when used on cells, but it is very effective on virus particles and on isolated DNA.

Nitrous acid reacts with amino groups to form hydroxyl groups according to the following equation:

$$R\text{–}NH_2 + HNO_2 \rightarrow R\text{–}OH + N_2 + H_2O$$

Thus, guanine can be deaminated to xanthine, adenine to hypoxanthine and cytosine to uracil. While the first reaction may be nonmutagenic (or lethal), the other two should lead to transition mutations, since hypoxanthine has the hydrogen-bonding properties of guanine and uracil has those of thymine.

Base Analogs Base analogs are incorporated into DNA in place of bases whose structure they mimic. Two analogs have been used extensively as mutagens for bacteria, namely 5-*b*romo*u*racil (5-BU), a thymine analog, and 2-*a*mino*p*urine (2-AP), an adenine analog. 5-BU, because of the electronegativity of the bromide atom which replaces the thymine methyl group, has a much greater tendency than has thymine to exist in the *enol* tautomeric form. 2-AP can probably use its 2-amino group to pair with cytosine as well as with thymine. Both base analogs induce transitions in both directions, as expected.

Acridine Compounds Proflavine, a member of the acridine family of compounds, was the first agent shown to induce frameshift mutations. This special mutagenic effect of acridine compounds is plausibly explained by the known ability of the acridine ring system, which mimics a hydrogen-bonded DNA basepair in shape and

hydrophobicity, to intercalate into the stack of basepairs running up the core of the DNA double helix (see also above).

5.2.3 Detection Systems for Mutations

Several systems have been developed allowing the detection of mutations in bacteria, among them:

- the Ames test
- the LacI test.

While the first one has been developed with *S. typhimurium*, the latter makes use of *E. coli* cells. Because any mutation assay system is based on the selection or screening of mutants, it is impossible to collect all the sequence alterations that occur in any given target gene. In addition, the kinds of mutations that can be isolated and the number of sites that can be detected depend on the genetic system used. The rate of mutation per locus differs by up to two orders of magnitude among target sequences. Furthermore, the target sequences differ in the number of mutational hotspot sites; and the strength of each hotspot can also vary. These observations make it difficult to collect weaker or nonhotspot types of mutations. In summary, relatively rare types of spontaneous mutations can easily be overlooked unless large numbers of independent mutations are examined. Base substitution mutations is the most frequent class of spontaneous point mutation, followed by single-base frameshifts, about two orders of magnitude less frequent.

The Ames Test

The most widely used test system to detect the mutagenic effect of chemicals was developed by Bruce Ames. It makes use of *S. typhimurium* strains carrying *his* point mutations which prevent growth on minimal medium in the absence of added histidine. If these strains are plated on such medium, a few cells can form colonies at a frequency of about 10^{-6} due to the spontaneous reversion of these point mutations. When these cells are treated with a chemical causing mutations, they will increase the spontaneous mutation rate. Therefore, the principle of the Ames test relies on the significant increase in the spontaneous mutation rate caused by the mutagen.

The Ames test is made more sensitive by one or more genetic changes in the original strain: (a) the introduction of a point mutation within the *uvrA* gene, thereby eliminating the nucleotide excision repair system (see below), (b) introduction of the pKM101 plasmid coding for the error-prone DNA polymerase V encoded by the *mucA* and *mucB* genes, (c) mutations in genes coding for outer membrane proteins, increasing the uptake of some mutagenic compounds. As mentioned above, some chemicals are not mutagenic themselves but can be converted into mutagens by enzymes in the mammalian liver. To obtain this conversion, a rat liver extract called S9 is used (the liver is homogenized and centrifuged

at 9000 *g*; and the supernatant is then mixed with the *Salmonella* test cells). The disadvantage of the Ames test is that only true point mutations can be detected, but not frameshift, deletion and insertion mutations. This led to the development of an additional test system, the LacI system.

The LacI Test

The *lacI* gene encodes the repressor of the *lac* operon; and *lacI*$^-$ mutants (carrying a defective repressor) express the *lac* operon constitutively. These mutants can be readily selected on media containing the noninducing sugar phenyl-β-D-galactoside as sole carbon source. Since no constraints are placed on the specific nature of the mutations, the system is useful in defining the specificity of mutations. One limitation of the *lacI* gene was the presence of a large hotspot (comprising about 70% of the mutations) that dominated the spectrum. This hotspot has been eliminated by the introduction of two silent base substitutions. Among 414 independent mutants, 70.8% turned out to be base substitutions, 17.2% deletions, 7.7% insertions and 4.3% single-base frameshifts.

5.2.4
Adaptive Mutations

When populations of bacteria are exposed to nonlethal selections, the nonproliferating cells accumulate mutations that relieve the selective pressure, a phenomenon called *stationary phase* or *adaptive mutations*. These selected mutations are accompanied by nonselective mutations. Most research on adaptive mutations has focused on an *E. coli* strain called FC40 that cannot utilize lactose, due to a frameshift mutation in *lacZ*, but that readily reverts to lactose utilization when lactose is its sole carbon source. The underlying process that produces Lac$^+$ mutations is not the same as that which produces Lac$^+$ mutations during normal growth. Two mechanisms of adaptive genetic change are now known to occur in the *lac* frameshift system (see below): the *adaptive point mutation* and the *adaptive gene amplification* mechanism. On selective medium, rare preexisting cells with a *lac* duplication initiate slow-growing clones within which the growth rate increases progressively as amplification enhances the copy number of the partially functional mutant *lac* allele. Adaptive point mutations and amplification are separate adaptive responses and are both different from Lac$^+$ mutation in growing cells.

E. coli Strain FC40

E. coli strain FC40 cannot utilize lactose (Lac$^-$) but readily reverts to lactose utilization (Lac$^+$) when lactose is its sole carbon and energy source. Two days after plating FC40 cells on minimal lactose plates, Lac$^+$ colonies start appearing at a constant rate of nearly one per 107 cells per day. Of the Lac$^+$ colonies that appear during a week of incubation on lactose, 95% are due to mutations that occur after plating. FC40 has been so extensively studied that it became a paradigm of adap-

tive mutations. The Lac$^-$ allele carried by FC40 is derived from a fusion of the *lacI* gene to the *lacZ* gene that eliminates the coding sequence for the last four residues of *lacI*, all of *lacP* and *lacO* and the first 23 residues of *lacZ*. Constitutive transcription is initiated from the *lacI*^{q} promoter. The Lac$^-$ allele carried by FC40, Φ(*lacI33-lacZ*), has an ICR191-induced +1 frameshift at the 320th codon of *lacI*, changing CCC to CCCC. The allele is slightly leaky, producing about 2 Miller units of β-galactosidase, which is not sufficient to allow FC40 to grow on lactose. The frameshift is polar on *lacY*; and, thus, FC40 is also lactose permease-defective. However, stationary-phase cells of FC40 do not revert to Lac$^+$ unless lactose is present, so energy provided by the residual amount of β-galactosidase and permease activity might be required for the mutational process.

The Adaptive Point Mutation Mechanism

The adaptive point mutation mechanism includes the following components: DNA breakage, recombinational break repair, transient mismatch repair limitation, genome-wide hypermutation in a subpopulation of cells and the mutator DNA polymerase Pol IV. The first step is the generation of a double-stranded end as a loading site for the RecBCD enzyme. Such a double-stranded end can arise when the replication fork arrives at a single-stranded nick at the *oriT* of the F′ factor or at any nick on the plasmid. Alternatively, a stalled and regressed replication fork can provide a loading site. In the second step, the double-stranded end is processed by RecBCD (see Fig. 4.1), creating a 3′-ended single-stranded DNA onto which RecA will load. The RecA nucleofilament catalyzes invasion of a homologous duplex to produce a displacement loop and leads to induction of the SOS response. The 3′ end is recognized by the error-prone Pol IV (induced by the SOS response) and the errors produced by this enzyme remain uncorrected due to a transient deficiency in methyl-directed mismatch repair (MutL is limiting; overproduction of MutL decreases adaptive mutations). Finally, the Holliday structure is removed by the RuvABC resolvasome. The idea of a hypermutable subpopulation has been confirmed by the analysis of nonselected secondary mutations. When Lac$^+$ cells were analyzed for such mutations, they were found to have unselected mutations throughout the genome (both F′ factor and chromosome) about two orders of magnitude higher than in the main population of Lac$^-$ cells. The hypermutable state is transient and has been estimated to be between 10^{-3} and 10^{-4} of all cells. What is the origin of the hypermutable subpopulation? One possibility is that it is SOS induction that separates the subpopulation since a starvation-induced SOS response has been described.

The Adaptive Gene Amplification Mechanism

A second, fundamentally different mode of formation of Lac$^+$ cells occurs by adaptive amplification. In adaptive amplification, the leaky *lac* mutant gene can be amplified to 20–50 copies of 7-kb to 40-kb direct DNA repeats, providing sufficient β-galactosidase activity for growth on lactose medium without acquisition of the

Lac^+ point mutation. Amplification is a more flexible genomic alteration than point mutations. It is a reversible process and can allow evolution of a gene copy while an intact copy of the gene is retained. It can be assumed that adaptive amplification is an important mode of adaptive genetic change in evolution and development. Cells carrying the amplified gene are not hypermutated in unselected genes. Though the underlying mechanism leading to amplification is unknown, the enzymes involved in homologous recombination seem to be needed, since totally adaptive Lac^+ colonies do not appear in their absence. This suggests that double-stranded breaks are involved which are processed in a nonhomologous recombination event that starts amplification.

P.L. Foster **1999**, Mechanisms of stationary phase mutation: a decade of adaptive mutation, *Annu. Rev. Genet.* 33, 57–88.
S.M. Rosenberg **2001**, Evolving responsively: adaptive mutations, *Nat. Rev. Genet.* 2, 504–515.

5.3
Repair of DNA Lesions

A basic feature of living organisms is their ability to repair their genomes. Cell are exposed to a multitude of DNA-damaging agents throughout their lifespan; and it is absolutely necessary for an organism to deal with these challenges effectively. Under normal conditions, there is no need to maintain a high level of repair and damage-processing activities inside the cell, but, after exposure to DNA-damaging agents, a cell should be able to respond promptly to process the DNA damage. To achieve this goal, organisms have developed elaborate systems and networks for the coordinated expression of genes required for the repair of DNA damage.

Because most mutations are detrimental, organisms have evolved repair mechanisms, sometimes also called *mutation avoidance mechanisms*, to keep their mutation rates as low as possible, given the costs, in terms of energy and time, of error prevention and repair. Mutations inactivating mutation-avoidance genes (mutator mutations) will lead to a permanent increase in the cellular mutation rate and are called constitutive mutators. Four major types of DNA repair mechanisms eliminate *premutagenic damage* (produced by a variety of factors present in normally growing cells) and *premutations* (the mutagenic intermediate):

- the proofreading function of the Pol III
- the mismatch repair system
- the nucleotide excision repair system
- various pathways of base excision repair.

The latter two mechanisms act mostly in the prereplicative elimination of endogenous DNA lesions; and both repair mechanisms cooperate in preventing mutagenesis caused by the miscoding and replication-blocking types of spontaneous DNA lesions. The principle of all three mechanisms of repair involves splicing out the damaged region and inserting new bases to fill the gap, followed by ligation of the pieces.

5.3.1
The Methyl-mediated Mismatch Repair System

The *m*ethyl-mediated *m*ismatch *r*epair system (MMR) plays a central role in the post-replicative correction that removes premutations before the next round of DNA replication. This repair system removes mismatches, single-base bulges and eliminates even the multiple-base mismatches that lead to sequence substitution mutations, but their repair largely depends on the size of the mismatched segment. Incorrect nucleotides are removed by a strand-specific excision reaction that is directed to daughter DNA strands by virtue of the transient absence of d(GATC) methylation within newly synthesized DNA.

Repair is initiated by a MutS homodimer (monomer molecular mass is about 90 kDa) binding to the mismatched bases or deletion/insertion loops which are one to four nucleotides in length and binding to seven out of eight possible mismatches, C:C being refractory (Fig. 5.2). Each MutS protein has a region that binds and hydrolyzes ATP, producing ADP. It seems that these regions also differ between the two MutS proteins. The ATP-binding domains interact with each other and with the DNA-binding regions of both proteins, allowing the effects of ATP/ADP binding and DNA binding to be transferred across the protein. These details provide a structural basis for the biochemical observation that the binding and hydrolysis of ATP results in conformational changes in the MutS protein and in the modulation of protein–DNA interactions. An intriguing question is how the MutS proteins bind specifically to DNA containing mismatched bases. These proteins have only about a 20-fold higher affinity for mispaired bases relative to normal base pairs; and mispaired bases are not all that common – they occur probably only once in every one million to ten million base pairs. The new structures suggest one possible answer: that only binding to a mispair induces the conformational changes seen in both MutS and the bound DNA. If these conformational changes are required for the MutS proteins to interact with other proteins

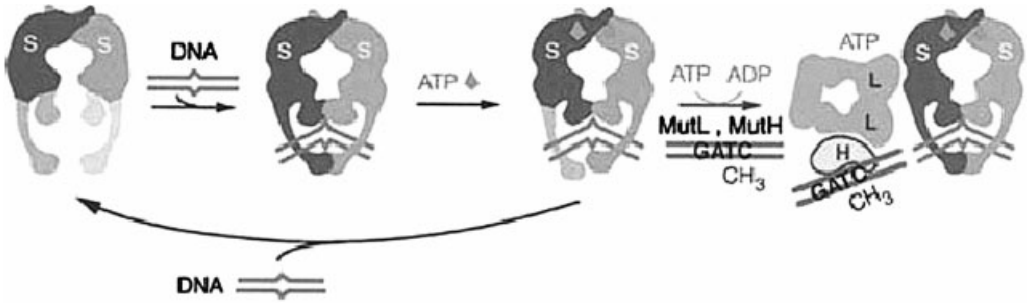


Fig. 5.2 Methyl-mediated mismatch repair. MutS recognizes and binds to the mismatch followed by binding of ATP. This complex recruits MutL which recognizes the nearest hemimethylated GATC and binds MutH. MutH cuts 5′ or 3′ to the unmodified GATC sequence, UvrD binds to the nick and unwinds the double-strand DNA followed by resynthesis and ligation. T.A. Kunkel, D.A. Erie **2005**, *Annu. Rev. Biochem.* 74, 681; Fig. 2C. (This figure also appears with the oclor plates.)

during mismatch repair, then these changes and subsequent protein–protein interactions may effectively increase mispair specificity. MutL, also a dimer, then binds the MutS-DNA complex, leading to the formation of a ternary complex and the DNA is looped out in an active search for the nearest d(GATC) methylation site, either 5′ or 3′ of the mismatch. Once found, the ternary complex activates the latent endonuclease activity of MutH, which incises the unmethylated (or nascent) strand at a hemimethylated d(GATC) site. MutH recognizes the asymmetry of the unmethylated N6 position of the d(A), and cleaves 5′ to the d(G) in the unmethylated strand. If both strands are unmethylated, MutH is able to cleave both sites independently [dissociating after the first d(GATC) cut, with a second binding and cleavage event], leaving a four-base overhang. In addition, this complex activates the unwinding activity of DNA helicase II (*uvrD*/*mutU* gene product), which loads at the MutH incision with an orientation bias, so that helix unwinding proceeds towards the mismatch. That portion of the unwound, incised strand is subject to degradation by one of several single-stranded DNA exonucleases; and DNA removed in this manner is resynthesized by Pol III holoenzyme in the presence of single-stranded DNA-binding protein. Finally, DNA ligase restores covalent continuity to the repaired strand. It is also important to note that the d(GACT) site nearest an error may be >1 kb away. The MMR system increases the overall fidelity of DNA synthesis by as much as 1000-fold.

5.3.2
Very Short Patch Repair

A mismatched basepair T:G can arise in DNA from misincorporation during replication or, to a greater extend, from the spontaneous hydrolytic deamination of 5-methyldeoxycytosine ($d^{5Me}C$) in existing basepairs. In bacteria, $d^{5Me}C$ is introduced into DNA by the post-replicative action of DNA-dC methyltransferases, which may function as one component of a restriction-modification system or may be a protein of uncertain function, such as the Dcm methyltransferase from *E. coli*. To counteract the mutagenic potential of T:G mismatches, arising from deamination of $d^{5Me}C$:G basepairs, many bacteria possess a ''very short patch'' mismatch repair (VSR) pathway characterized by the presence of a ''very short patch repair'' endonuclease, the product of the *vsr* gene. The Vsr endonuclease cleaves the phosphodiester bond 5′ to the mismatched T within the sequence 5′-C↓T(A/T)GG3′/5′CC(T/A)GG-3′ (Fig. 5.3) producing a nick containing 5′ phosphate and 3′ hydroxyl termini. Repair is completed by the 5′→3′ exonuclease and DNA polymerase activities of Pol I and DNA ligase. The exonuclease of Pol I excises only a short stretch, approximately five residues, before resynthesis and ligation of the corrected strand. Methylated cytosines are vulnerable to spontaneous deamination and hydrolysis, which result in the formation of a mismatched thymine. Therefore, the *vsr* genes are invariably associated with a gene coding for a DNA dC-methyltransferase; and thus, in *E. coli*, the *vsr* and *dcm* genes overlap and ensure concerted transfer during transduction and conjugation.

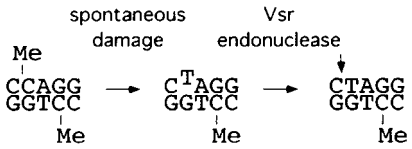


Fig. 5.3 Very short patch repair. *E. coli* methylates the internal C residues in the sequence 5′-CCWGG-3′ catalyzed by the Dcm methylase. Spontaneous deamination converts the 5-methylcytosine to thymine, a normal base, which mispairs with guanine, resulting in a transition mutation after one round of DNA replication. The Vsr endonuclease recognizes specifically T:G mismatches and cleaves 5′ to the mismatched T. Pol I binds to the nicks, removes a few nucleotides and replaces them. DNA ligase then seals the final nick. M. Lieb **1996**, *Mol. Microbiol.* 20, 467–473; Fig. 1.

5.3.3
The Nucleotide Excision Repair Systems

Among the most highly conserved biochemical pathways in free-living organisms are those involved in DNA repair. The ubiquitous pathway of *n*ucleotide *e*xcision *r*epair (NER) is responsible for the removal of environmentally induced DNA damage, such as the DNA lesions resulting from sunlight exposure or chemical carcinogens. The study of DNA repair in *E. coli* has helped us to understand the corresponding repair pathways in humans. NER can be viewed in four basic steps:

- damage recognition and lesion verification (UvrA, UvrB)
- incision (UvrC)
- excision (UvrD)
- repair synthesis and ligation (Pol I, DNA ligase).

Damage recognition and verification are achieved by a protein machine that utilizes several components to sense distortions in the double-helical duplex DNA. In *E. coli*, the UvrA and UvrB proteins, both forming homodimers, carry out these functions (Fig. 5.4). If the putative lesion is identified by UvrA, the repair complex enlists the strand-opening activity of UvrB that helps to verify that the distortion is, in fact, due to a damaged nucleotide. It is believed that the β-hairpin domain of UvrB, rich in hydrophobic residues, is inserted into the DNA helix both to verify the damaged nucleotide and to establish which strand has been damaged. When the UvrB protein has verified the presence of the damage, UvrA dissociates to leave the UvrB-DNA complex. During the damage recognition process, the DNA is wrapped around the UvrB protein and this DNA wrap is dependent on ATP binding by UvrB. DNA wrapping is expected to cause local melting of the DNA helix, thereby facilitating insertion of the β-hairpin of UvrB between the DNA strands. From mutational studies of the β-hairpin, it was proposed that UvrB scans the DNA for damage by trying to flip nucleotides out of the DNA helix, thereby probing for differences in base stacking. When no damage is present, the nondamaged nucleotide will be held in place by its neighbors. Only when damage is detected can the UvrB-DNA preincision complex be formed. This complex is

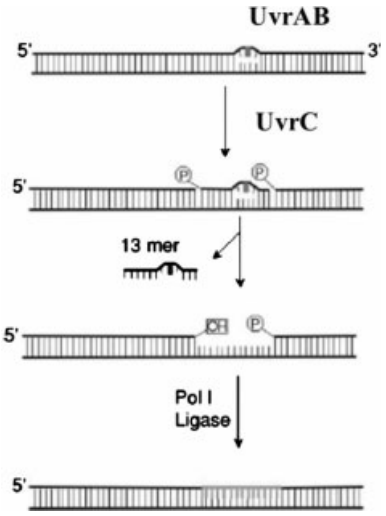


Fig. 5.4 Nucleotide excision repair. The UvrAB complex identifies lesions, UvrC cuts the damaged strand four to five nucleotides 3′ to the altered nucleotide and about eight nucleotides 5′ from the damaged nucleotide. The UvrD helicase removes the strand with the damaged nucleotide, and the gap is filled in by Pol I and sealed by the DNA ligase. A. Sancar, J.T. Reardon **2004**, *Adv. Protein Chem.* 69, 43; Fig. 1, modified.

subsequently bound by the UvrC protein, which catalyzes both incisions. The UvrC protein consists of two functional parts. The N-terminal part of this protein is responsible for cutting the damaged strand four or five nucleotides 3′ to the altered nucleotide, while the C-terminal part is necessary and sufficient to produce the second incision, some eight nucleotides 5′ from the damaged nucleotide, at about one turn of the helix from the first cut. Once incision has occurred, the damaged oligonucleotide is excised by the UvrD helicase, just before or during the repair synthesis step, which is normally carried out by Pol I. DNA ligase then seals the newly created repair patch at its 3′ end to complete the process.

NER can repair essentially any bulky lesion by cutting out a 12-base single-stranded region containing the lesion; but it fails to repair, or repairs very poorly, small lesions such as methylated and oxidized bases.

Recently, an alternative NER pathway has been described which is assumed to be used for some types of bulky DNA adducts that may interfer with UvrC incision. In this pathway, UvrC is replaced by the Cho (Uvr*C ho*mologue) protein that is homologous to the N-terminal part of UvrC. Cho only produces incisions on the 3′ side of the lesion, at the ninth phosphodiester bond 3′ to the altered nucleotide, four nucleotides beyond the site of the normal UvrC incision. Cho requires the formation of the UvrB-DNA damage verification complex but does not require the UvrC binding domain of UvrB. How important is Cho? While Cho and UvrC coexist in some bacterial species (*E. coli*, *Clostridium*, *Listeria*, *M. tuberculosis*), many more species only have UvrC, while Mycoplasmas and *B. burgdorferi* have only Cho. The *Mycobacterium* Cho protein has an additional domain with a strong homology to the proofreading subunit of Pol III, suggesting that, once the 3′ incision is produced, the putative exonuclease activity of this domain would digest in the 3′ direction through the lesion site and leave a 3′ OH end that could serve as a primer for repair synthesis. What could be the functional role of Cho in those bacterial species which contain already UvrC? Some types of bulky DNA adducts may

interfere with the nearby UvrC incision while Cho, cutting four nucleotides farther away from the lesion, would not be inhibited.

5.3.4 Pathways of Base Excision Repair

Base excision repair (BER) is a multiprotein pathway that differs from nucleotide excision and mismatch repair in that the broad substrate specificity is determined by diverse damage-specific DNA glycosylases. They act on a wide range of DNA adducts that result from the action of a number of DNA-damaging agents including water, reactive oxygen species and alkylating agents. The modified bases that are removed in this way include the products of hydrolytic deamination of cytosine to uracil, 5-methylcytosine to thymine, oxidative and alkylation damage products. While DNA glycosylases such as uracil DNA glycosylase are very specific regarding their substrate, others, such as endonuclease III, can excise a wide variety of damaged DNA bases.

The glycosylases are relatively small monomeric proteins that process damaged bases with high specificity. In the first step of the BER pathway, a lesion-specific DNA *N*-glycosylase removes the damaged base by cleaving the glycosidic bond between the base and deoxyribose of the DNA backbone (Fig. 5.5). Base excision glycosylases locate damaged bases within a vast excess of unmodified DNA; and they expose a target nucleotide for cleavage of the glycosylic bond by a reaction that does not require a high energy cofactor. We know little about how DNA glycosylases search for DNA damage or about the mechanism and energetics of distorting DNA structure to capture the modified nucleotide in the active site of the glycosylase. Crystal structures of DNA glycosylases complexed to DNA substrates have revealed that an unpaired nucleotide of the DNA substrate fits into a pocket on the enzyme surface by a process termed base-flipping or nucleotide-flipping. Damaged bases might be more susceptible than normal bases to being flipped out of the DNA helix and into the glycosylase active site. A subset of these enzymes has an associated *ap*urinic/*ap*yrimidinic (AP) lyase activity that, via a β-elimination reaction, produces 3′ αβ-unsaturated aldehyde and 5′ phosphate products. The glycosylase activity leads to the production of a common intermediate in the pathway, the AP site. These sites are subsequently processed by AP endonucleases or AP lyases that cleave the phosphodiester bond either 5′ or 3′ to the AP site, respectively. This site is processed further to generate a 3′ OH end used in polymerization by Pol I and ligation.

Repair of Alkylation Damage

E. coli expresses two BER glycosylases that excise alkylation-damaged bases from DNA. The constitutively expressed *E. coli* Tag protein (3-methyladenine DNA glycosylase I) is relatively specific for the removal of 3-methyladenine, whereas the inducible AlkA protein (3-methyladenine DNA glycosylase II; see also below) removes a structurally diverse group of alkylated bases from DNA. The crystal struc-

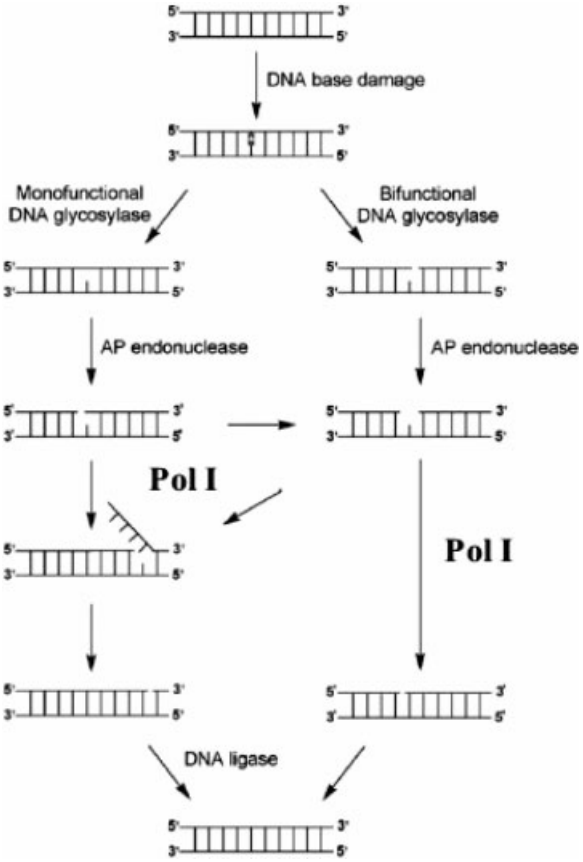


Fig. 5.5 Base excision repair. A lesion-specific DNA glycosylase recognizes and removes the damaged base, leading to an AP (*ap*urinic/*apy*rimidinic) site. DNA glycosylases fall into two mechanistic classes: monofunctional DNA glycosylases displace the lesion by using a molecule of solvent water, while bifunctional DNA glycosylases displace the lesion by using a nucleophilic active site on the enzyme. Next, an AP endonuclease binds to the AP site and removes a few nucleotides. The resulting gap is filled in by Pol I and sealed by DNA ligase. J.C. Fromme, G.L. Verdine **2004**, *Adv. Protein Chem.* 69, 1; Fig. 1, modified.

ture of AlkA reveales a broad cleft located between two protein subdomains that contains a catalytically essential aspartic acid residue. The 2.5-Å crystal structure of AlkA complexed to DNA showed that the enzyme flips a 1-azaribose abasic nucleotide out of DNA and induces a 66° bend in the DNA, with a marked widening of the minor groove.

Uracil and Hypoxanthin Repair

The RNA base uracil arises in DNA either by erraneous incorporation by DNA polymerases or as a consequence of cytosine deamination by physiological or en-

vironmental factors. If left unrepaired, in a U:G mismatch, uracil is premutagenic and can lead to GC→AT transition mutations during the next round of replication. Further, uracil in an A:U pair in the DNA sequences can impede their recognition by the cognate regulatory proteins. In cells, a highly efficient base excision repair enzyme, *u*racil-*D*NA *g*lycosylase (UDG), is involved in freeing the DNA from uracil residues. UDGs act by cleaving the glycosidic bond between uracil and the sugar by the attack of a hydroxyl nucleophile on the desoxyribose C1′ atom. This nucleophile is generated by the activation of a water molecule by an absolutely conserved Asp of the GQDPYH motif. Concomitant protonation of the O2′ of the uracil base by His of yet another highly conserved motif, HPSPLS, enhances its leaving group quality. How can phages survive containing uracil residues in their genome? In the case of the *B. subtilis* phages PBS-1 and PBS-1, a small 84-amino-acid protein (9.4 kDa; called Ugi for *u*racil DNA *g*lysosylase *i*nhibitor) was detected forming an extremely specific and exclusively stable complex with the host UDG in 1:1 stoichiometry.

5.3.5 Repair of Oxidized Nucleotides

While molecular oxygen (O_2) does not cause any harm to DNA and other macromolecules, reactive oxygen species have more electrons than molecular oxygen and are considered to be a major threat to the integrity of DNA, as well as proteins, lipids and carbohydrates. In aerobically growing cells, reactive oxygen species are produced as byproducts of normal metabolic pathways. Some of these byproducts include singlet oxygen ($^1O_2$), peroxide radicals ($\bullet O_2$), hydrogen peroxide (H_2O_2) and hydroxyl radicals (•OH). Although H_2O_2 is relatively stable, it can rapidly react with Fe^{2+} to produce highly reactive •OH radicals in a process described by the Fenton reaction. This •OH radical can then react with DNA to produce a variety of DNA lesions. Hydrogen peroxide is probably formed inadvertantly by flavoenzymes during oxidative respiration. Alternatively, reactive oxygen species can result from environmental factors, including UV irradiation and chemicals such as the herbicide paraquat (see also Section 9.5).

Repair of Oxidized Pyrimidines

The •OH radical causes the formation of saturated bases such as glycol, a replication blocking compound. Oxidized pyrimidine residues such as thymine glycols, 5,6-dihydrothymine, 5-hydroxy-2′-deoxycytidine and other pyrimidine radiolysis products are recognized by the *E. coli* endonuclease III (*nth* gene). This 23.5-kDa protein is an iron-sulfur cluster enzyme with both glycosylase and AP lyase activity. In addition, there is a second glycosylase/AP lyase enzyme specific for oxidized pyrimidines, endonuclease VIII, encoded by the *nei* gene. This endonuclease incises DNA containing thymine glycol, dihydrothymine and AP sites, as well as oxidative products of cytosine such as 5-hydroxycytosine and 5-hydroxyuracil. The X-ray structure of the endonuclease III has been determined at 1.85 Å re-

solution. It is an elongated bilobal structure with a deep cleft separating two distinct domains linked by a flexible linker region. The two domains are α-helical in structure and the C-terminus contains a 4Fe4S cluster believed to stabilize the structure of the enzyme. The 4Fe4S cluster coordinated by four cysteine residues has been implicated as a domain for DNA recognition and specific DNA binding, based primarily on the surrounding positive residues. A second novel binding motif within this enzyme is the helix-hairpin-helix (HhH) motif. This motif has been identified within 14 homologous families of proteins that include AlkA, MutY and the β subunit of the RNA polymerase. It has been suggested that this motif may be a conserved domain for nonsequence-specific DNA binding. The substrate binding site was identified as a β-hairpin structure. The structure identifies the binding site and catalytic residues within a pocket in the enzyme and suggests that the damaged base is flipped extrahelically into this active site pocket.

Repair of Oxidized Purines

A predominant lesion found in DNA exposed to reactive oxygen species is 7,8-dihydro-8-oxoguanine (8-oxoG or GO; Fig. 5.6A) which is produced by an •OH attack on C-8 of guanine. This oxidation product is especially deleterious due to its ability to form a stable Hoogsten base pair with adenine in addition to the canonical Watson–Crick base pair with cytosine. The facile misincorporation by DNA polymerase of an adenine opposite a GO results in a premutagenic adenine:oxoG mismatch, a site where further replication prior to repair would lead to C→A or G→T transversion.

E. coli cells have evolved a mechanism involving base excision repair that limits mutagenesis as a result either of the formation or incorporation of GO in DNA or of the replication of DNAs contaminating 8-oxoG lesions involving different *mut* genes. During replication, dAMP can be selectively incorporated opposite GO (Fig. 5.6B). The misincorporation of an adenine reflects the propensity of GO to adopt a *syn* conformation about the glycosidic bond that allows base pairing with adenine. There are three lines of defense to prevent the mutagenic effect of GO (Fig. 5.6B):

1. To remove a mispaired A opposite a GO: the adenine-specific DNA *glycosylase MutY* is responsible for the removal of the mispaired adenine in an effort to avoid the G→T transversion mutation. Besides the A:8-oxoG mispair, MutY removes adenine from A:G, A:C and A:8-oxoA in an effort to prevent the formation of C:G to A:T transversions. After removal of the adenine, a gap is formed which is repaired by Pol I leading to the incorporation of a cytosine. Here, the GO is still present in the DNA.

2. To excise a GO from the DNA: GO is removed by the specific *glycosylase MutM*. Next, the depurinated strand is cleaved by the AP endonuclease activity of MutM itself, followed by degradation by an exonuclease; and finally, the gap is resynthesized by Pol I. The *mutM* gene is induced in cells that have accumulated reactive oxygen species because it is part of the regulon (see Section 9.5).

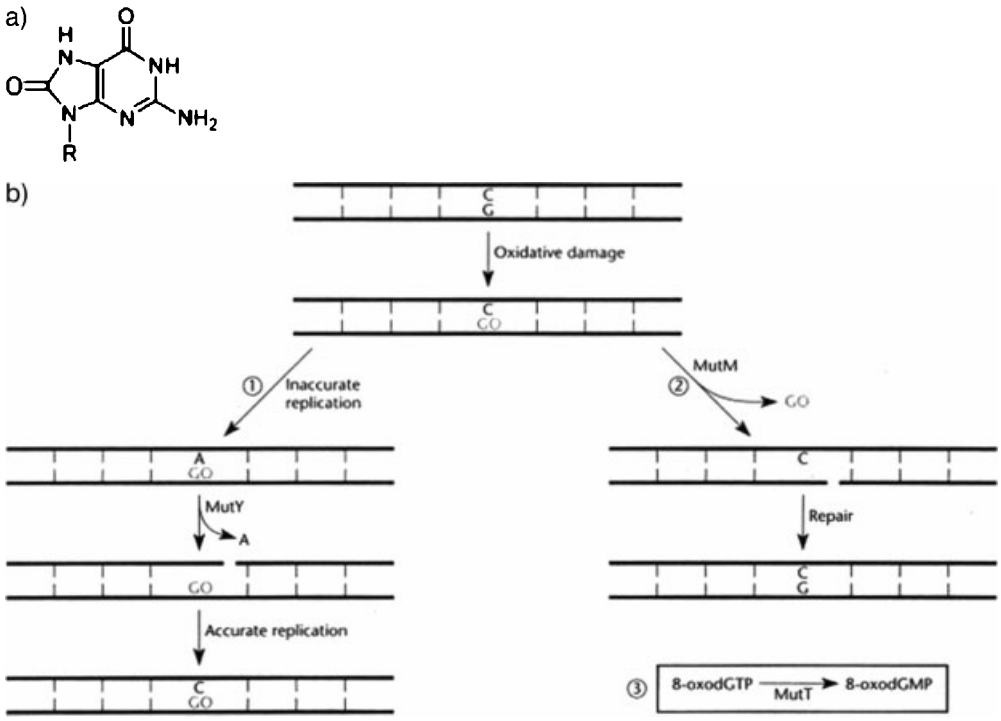


Fig. 5.6 Repair of oxidized purines. (A) Structure of 8-OxodG. (B) The MutY glycosylase removes a mispaired A opposite a GO (1); the glycosylase MutM removes the GO (2), and the MutT phosphatase specifically degrades 8-oxodGTP to prevent its incorporation into the DNA (3).

3. To prevent incorporation of 8-oxodGTP into DNA: the reactive forms of oxygen can oxidize not only guanine in DNA to GO, but also the base in dGTP to form 8-oxodGTP. This oxidized form is recognized by the *MutT* phosphatase, that specifically degrades 8-oxodGTP to 8-oxodGMP to prevent its incorporation into DNA.

5.3.6 Photoreactivation of Cyclobutane Dimers

UV irradiation is the major source of natural damage due to sun exposure. The conjugated ring structure of the bases in the DNA causes them to strongly absorb light in the UV wavelength. The absorbed photons energize the bases; and their double bonds react with nearby atoms, leading to the formation of additional chemical bonds. These chemical bonds can cause a linkage between the bases in the DNA and other bases or between bases and the sugars of the nucleotide. One common type of UV irradiation damage is the formation of pyrimidine dimers, in which the rings of two adjacent pyrimidines become fused. Two possible dimers can be formed. When the 5- and 6-carbon atoms of two adjacent pyrimidines are

joined, a cyclobutane ring is formed. In the other type, the 6-carbon of one pyrimidine is joined to the 4-carbon of an adjacent pyrimidine to form a 6-4 product.

Photoreactivation is a special type of repair system designed to resolve the cyclobutane-type pyrimidine dimers. This repair system separates the fused bases and depends on the enzyme *photolyase* encoded by the *phr* gene. This enzyme contains a reduced flavin adenine dinucleotide ($FADH_2$) group that absorbs light at wavelengths between 350 nm and 500 nm. It binds to the fused bases and the absorbed light provides the enzyme with the energy necessary to separate the fused bases. It follows that photoreactivation occurs only in the presence of visible light.

5.3.7
Transcription-coupled Nucleotide Excision Repair

What happens when the elongating RNA polymerase arrives at a noncoding region with a DNA lesion such as a cyclobutane dimer? The stalled polymerase is recognized by the Mfd (for *m*utation *f*requency *d*ecline) enzyme expressing two activities:

- It recruits the DNA excision machinery.
- It releases the RNA polymerase and its transcript from the DNA.

The *E. coli* Mfd is a 130-kDa multifunctional protein with three identified functional domains: Near its N-terminus, a UvrB homology segment is believed to recruit the excision repair complex to the stalled RNA polymerase followed by a RID (*R*NA polymerase *i*nteracting *d*omain) which binds Mfd to the polymerase. Near its C-terminus, we find an element assumed to translocate RNA polymerase along the DNA, leading to its release. Upon binding to the stalled RNA polymerase, Mfd can translocate the polymerase backwards (backtrack) from the DNA lesion, promote its dissociation from the DNA and recruit the UvrBCD complex which repairs the lesion. Such a cellular mechanism restoring the template should be of clear value to the cell's survival.

J. Roberts, J.-S. Park **2004**, Mfd, the bacterial transcription repair coupling factor: translocation, repair and termination, *Curr. Opin. Microbiol.* 7, 120–125.

5.3.8
The Adaptive Response

Alkylating agents are ubiquitous in our environment and are classified as electrophiles that react with nucleophilic centers in DNA to generate alkylated lesions. While S_N2 alkylating agents alkylate DNA at base nitrogens, S_N1 agents alkylate at both nitrogens and oxygens. Alkyl-base lesions have the potential to be mutagenic, can halt replication and interrupt transcription.

The adaptive response to alkylation damage is a positively regulated response specifically induced by methylation damage to DNA and includes the following four genes: *ada*, *aidB*, *alkA* and *alkB*. The Ada (*ada*ptive) protein has two functions:

- It is a suicidal DNA methyltransferase that demethylates two methylated bases (O^6-methylguanosine and O^4-methylthymidine) and methylphosphotriesters produced by methylating agents in the sugar phosphate backbone.
- It acts in its methylated form as a transcriptional activator of its own gene and of the three genes *alkA*, *alkB* and *aidB* organized altogether in three transcriptional units.

The Ada protein has two independent methyltransferase activities able to repair DNA. One is located in its N-terminal domain, the second in its C-terminal domain (Fig. 5.7). The C-terminal methyltransferase removes methyl groups from O^6-methylguanosine and O^4-methylthymidine and transfers them to its own Cys321 residue, thereby restoring the bases to their undamaged state. The second methyltransferase activity is required for the removal of methyl groups from the phosphate oxygens, repairing only the Sp diastereoisomer and leaving the Rp diastereoisomer to remain in the DNA. The methyl groups are transferred to the Cys69 residue of Ada, thereby converting the protein into a transcriptional activator that stimulates transcription from its own promoter and from the *alkA* and *aidB* promoters. Since methylation of the Ada protein is irreversible, one Ada protein molecule can repair only one lesion with each of its two methyltransferase activities.

The mechanism of transcription activation by Ada is complex, beause the interaction between Ada and RNA polymerase at the *alkA* promoter appears to be different from that at the two other promoters. A striking feature of the *ada* and *aidB* promoters is the presence of UP elements, DNA sequences that are recognized by the αCTD of the α subunit of the RNA polymerase and function as enhancers of transcription (see Section 6.2.1). The *ada* and *aidB* UP elements function as αCTD binding sites and enhance basal and ^{me}Ada-dependent transcription. At

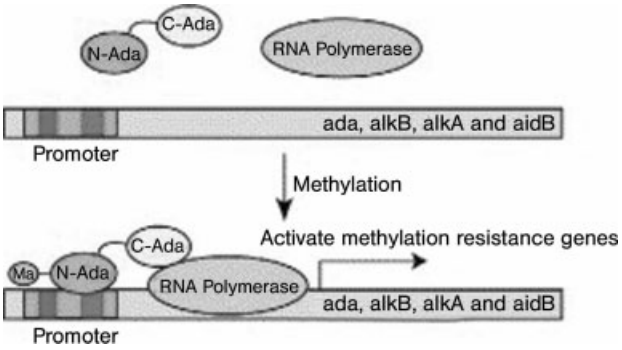


Fig. 5.7 Regulation of the adaptive response. The Ada regulon consists of four genes organized in three transcriptional units. The Ada protein, a transcriptional activator, is activated by transferring two methyl groups to two own cysteine residues. The activated Ada protein binds to Ada boxes preceding the promoters of the three transcriptional units thereby increasing transcription of the downstream genes. C. He et al. **2005**, *Mol. Cell* 20, 117; Fig. 1b.

the *alkA* promoter, RNA polymerase does not bind in the absence of the ^{me}Ada protein which contacts a positively charged patch in σ and also interacts with the α subunit.

AlkA is a 3-methyladenine-DNA glycosylase II that repairs several different methylated bases. The *alkB* gene, which forms a bicistronic operon with *ada* is required for error-free replication of methylated single-stranded DNA. 1-Methyladenine and 3-methylcytosine are the predominant forms of base damage only in single-stranded DNA, because these modification sites are normally protected by base pairing. These lesions can become cytotoxic because they stall the replisome. Purified AlkB protein is able to repair these cytotoxic lesions in single- and double-stranded DNA in a reaction that is dependent on oxygen, α-ketoglutarate and Fe(II). The enzyme couples the oxidative decarboxylation of α-ketoglutarate to the hydroxylation of these methylated bases in DNA, resulting in direct reversion to the unmodified base and the release of formaldehyde. The AidB protein has been suggested to detoxify nitrosoguanines, reducing the level of methlylation by these agents.

P. Landini, M.R. Volkert **2000**, Regulatory responses of the adaptive response to alkylation damage: a simple regulon with complex regulatory features, *J. Bacteriol.* 182, 6543-6549.

5.3.9
The SOS Response

At least 32 genes are known in *E. coli* to be under the negative control at the transcriptional level by the LexA repressor; and these genes form the LexA regulon. Upon damage of the DNA by UV radiation or chemicals such as mitomycin C, expression of the genes of the LexA regulon is increased; and this response has been designated the SOS response. The products of these genes function in DNA repair, recombination, mutation, translesion DNA synthesis and prevention of cell division. Induction of the LexA regulon needs the activation of the coprotease activity of the RecA protein, which facilitates autocleavage of the LexA repressor (and some other proteins), leading to the transient induction of genes of the LexA regulon. After repair of the DNA damage, the amount of activated RecA drops, leading to the increase of active LexA repressor by *de novo* synthesis, which in turn switches off all the genes of the LexA regulon. Members of the LexA regulon include LexA and RecA, the DNA polymerases II, IV and V, UvrABD, RuvAB and SSB. It should be mentioned that a number of genes have been identified as being induced by DNA damage, but not belonging to the LexA regulon, including *dnaA*, *dnaB*, *dnaN*, *dnaQ*, *recQ* and *phr*. They might be members of another regulon, the regulator of which has still to be identified.

Regulation of the SOS Response

In the absence of extensive DNA damage, the LexA repressor negatively regulates transcription of at least 32 genes of the LexA regulon by interacting with an opera-

Table 5.1 List of LexA-regulated genes and operons and their functions.

Genes and operons	*Function of the gene product(s)*
dinA (*polB*)	DNA polymerase II
dinB (*dinP*)	DNA polymerase IV; error-prone
dinD	Function unknown
dinF	Located downstream of *lexA*; function unknown
dinG	Putative helicase
dinI	Inactivates RecA* by protein-protein interaction
dinK	Function unknown
dinL	Function unknown
dinM	Function unknown
dinN	Function unknown
dinO	Function unknown
ftsK	Coordinates cell division with chromosome segregation
lexA	Negative regulator of the LexA regulon
recA	Involved in homologous recombination and induction of the LexA regulon
recN	Recombinational repair
ruvAB	Recognizes Holliday structures
sfiA (*sulA*)	Binds to the FtsZ protein and thereby prevents cell division
ssb	Single-stranded binding protein SSB
umuCD	DNA polymerase V; error-prone
uvrAB	Involved in nucleotide excision repair
uvrD	Involved in nucleotide excision repair

tor sequence (Table 5.1). Most interestingly, this operator sequence (20-bp palindromic consensus sequence: 5′-TACTGTATATATATACAGTA-3′) called the SOS box can vary in number and in location with respect to the promoter. While the *recN* gene is preceded by three, *lexA* itself and two plasmid-encoded genes (*cea*, *caa*) are controlled by two LexA operator sites. All the remaining genes contain just one operator either overlapping with the –35 region (*polB*, *uvrA*, *ruvAB*), or located between the –35 and the –10 region (*recA* and *uvrB*), or overlapping with the –10 region (*sulA*, *umuCD*) or downstream of the -10 region (*uvrD*). What could be the biological significance of these different locations? Could they reflect differential inducibility? Indeed, it has been shown that not all the genes are induced at the same time and induction depends on the severity of the DNA damage. As can be imagined, binding of the LexA repressor is determined by the primary structure and by the number of operators located in the promoter region. It should be mentioned here that the *umuCD* operon is the one with the tightest regulation; and this operon is induced only after severe DNA damage (see below). DNA damage induces the SOS response, which leads to the inactivation of the LexA protein by self-cleavage.

How is this accomplished? What is the nature of the SOS-inducing signal? Severe DNA damage inhibits DNA replication transiently, which leads to the accumulation of single-stranded DNA. Single-stranded DNA can be produced by a variety of mechanisms. First, single-stranded DNA breaks can be converted to single-stranded gaps by exonucleolytic digestion of the broken strand. Second, double-stranded DNA breaks can be converted to single-stranded regions by the activity of exonuclease V (see Section 4.1.4). Third, single-stranded DNA can be produced by replication of lesion-containing templates. And this single-stranded DNA provides the metabolic signal that activates the coprotease activity of RecA (designated RecA*). In its activated state, RecA* forms a spiral nucleoprotein filament on DNA (Fig. 5.8). Free LexA protein recognizes this structure and binds within the deep helical groove of the RecA* nucleoprotein filament. This interaction leads to the efficient autocatalytic cleavage of LexA at a scissile peptide bond between residues Ala-84 and Gly-85 within a hinge region of LexA that connects its N-terminal DNA binding and C-terminal dimerization domain. In this facilitated autodigestion mechanism, Ser-119 of LexA serves as the nucleophile that cleaves the Ala-84-Gly-85 bond. It is not clear how RecA* stimulates cleavage. One possibility is a conformational model in which RecA* stabilizes a reactive confor-

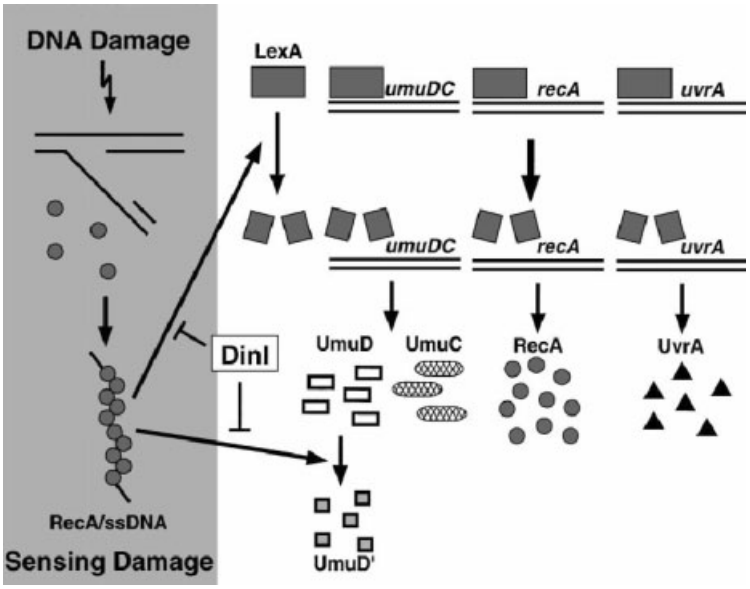


Fig. 5.8 Regulation of the SOS reponse and UmuD processing. Severe DNA damage leads to the appearance of ssDNA to which RecA will bind. Bound RecA will undergo a conformational change to an activated form, RecA*, acting as a coprotease for several proteins. One of them, the transcriptional repressor LexA will catalyze self-cleavage upon contact with RecA*, leading to the induction of all genes controlled by LexA, such as *umuCD*, *recA* and *uvrA*, called the SOS response. Another substrate is UmuD, which will self-catalyze into the active protein UmuD′. DinI, another member of the SOS regulon, inhibits the coprotease of RecA*. M.D. Sutton et al. **2000**, *Annu. Rev. Genet.* 34, 479; Fig. 1, modified.

mation of LexA. However, it also remains possible that RecA* makes a more direct contribution to the chemistry of bondage breakage. Cleavage of LexA reduces the level of functionally active LexA protein able to bind to its different operator sequences, leading to the induction of these genes. Both cleavage products of LexA are substrates for the ClpXP protease which uses sequence motifs that flank the auto-cleavage site but are dormant in intact LexA.

Much less is known about the events leading to the termination of the SOS response. Shut-down of the SOS machinery is an important step in the regulation of the SOS response: it ensures the resumption of DNA replication, reduces high mutability and reverses cell filamentation. It is widely believed that bacteria revert to the default (uninduced) state in a passive manner. That is, after DNA lesions are healed, the induction signal disappears and the original pool of LexA protein is restored. No active mechanism for turning down the SOS response has been described so far. But a negative regulator of the SOS response has been identified, termed DinI. This 81-amino acid protein (8.8 kDa) physically interacts with RecA* and prevents the binding of single-stranded DNA to RecA*, which is required for the activation of the latter. DinI also displaces single-stranded DNA from a stable RecA*-DNA cofilament, thus eliminating the SOS signal. When DinI is artificially overproduced, it prevents induction of the SOS response. Whether DinI only mitigates the SOS response or plays a more active role in shutting it down remains elusive. At the least, DinI exerts a fine-tuning role in the mutagenesis process by preventing extensive processing of UmuD.

There are at least two additional proteins which undergo self-cleavage when they encounter RecA*: the bacteriophage λ CI repressor and certain so-called mutagenesis proteins such as UmuD, MucA and RumA, where UmuD is encoded by the *E. coli* chromosome and MucA and RumA are encoded by plasmids. Key elements of the autocatalytic cleavage reaction include a conserved Ala(Cys)-Gly cleavage site and appropriately spaced Ser and Lys residues. Post-translational processing of UmuD generates the shorter, but mutagenically active UmuD′ protein (see below). Here, Ser-60 serves as the nucleophile that cleaves the UmuD Cys-24-Gly-25 bond to yield UmuD′. Similarly, the plasmid- and prophage λ-encoded proteins MucA, RumA and CI undergo self-cleavage upon contact with RecA*. In the case of MucA and RumA, the mutagenic proteins MucA′ and RumA′ are produced, acting in a similar way as described below for UmuD′, while cleavage of CI results in the induction of the resident prophage, thereby allowing λ to escape a cell which might be killed. Cleavage of UmuD and CI is slower than that of LexA, ensuring prophage induction and formation of the mutagenic protein only under conditions of severe DNA damage.

The *psiB* gene (144 aa) conserved among conjugative plasmids such as F is responsible for a plasmid-encoded anti-SOS function. It prevents not only induction of *sfiA* (=*sulA*) and prophage λ, but also mutagenesis and intrachromosomal recombination. *psiB* is located near *oriT*, the origin of conjugational transfer; and it undergoes zygotic induction. *psiB* is not under SOS control; and its role is believed to prevent the recipient cells from inducing the SOS response during the transfer of the single-stranded DNA.

Function of Genes Induced by the SOS Response

Most of the known genes of the LexA regulon are listed in Table 5.1. Many of them were identified in a special screen and are called *din* for *d*amage *in*ducible, while some others had already been identified before. The products of all these genes help the cell to survive DNA damage. While some encode products that are part of the nucleotide and recombination repair pathways, others transiently delay cell division (*sfiA*) until the damage can be repaired and others allow the cell to replicate past the DNA lesion. One of the hallmarks of the SOS response is that it is graded in such a way that error-free pathways of DNA repair are induced early (so that DNA fidelity of DNA repair remains high), whereas pathways that may ensure survival under more severe conditions (but are error-prone) are induced much later. Key participants in the latter pathway, which are often called SOS mutagenesis, are UmuD′ and UmuC proteins. We will discuss the role of the three SOS-inducible DNA polymerases and the mechanism of SOS mutagenesis in detail.

Pol II, the only SOS polymerase having a 3′ exonuclease proofreading activity, is induced ~7-fold from ~50 to 350 molecules per damaged cell almost immediately (~30 s) after SOS induction, whereas the delayed appearance of Pol V, ~30–45 min post-SOS induction, allows mutation-free repair processes to occur prior to error-prone translesion synthesis. In the absence of SOS induction, the three polymerases are present at about 50 (II), 250 (IV) and 15 (V) molecules/cell. The elevated level of expression in the absence of DNA lesions implies a role for Pol IV during normal DNA synthesis; and two different roles have been discussed. First, it has been suggested that Pol IV might help in rescuing stalled replication forks which are thought to collapse in the absence of DNA damage once per round of replication. These stalled replication forks might result from mismatched or misaligned primer ends refractory to proofreading by Pol III. By interacting with the β-processivity clamp, Pol IV might dissociate the Pol III core complex to extend aberrant primer ends. A second role for Pol IV has been discovered by genetic assays showing that this enzyme is primarily responsible for enhanced mutability in starving nondividing cells. This phenomenon has been designated ''adaptive mutation'' and is discussed in detail (see above).

The *umuCD* gene products participate in *two temporally separated* roles in response to DNA damage. First, the $UmuD_2C$ complex plays a noncatalytic role in promoting cell survival by participating in a DNA damage checkpoint control. This checkpoint acts to regulate the rate of DNA synthesis in response to DNA damage, thereby allowing additional time for accurate repair processes, such as nucleotide excision repair, to remove lesions in the DNA before the cell's attempts to replicate its genome. Approximately 25 min after irradiation with 25 J m^{-2} of UV light, UmuD undergoes a RecA*-facilitated autodigestion; and this post-translational modification serves to remove the N-terminal 24 residues of UmuD, yielding UmuD′. UmuD′ and UmuD are dimers, both in complex with UmuC and when isolated *in vitro*. They also form UmuD′-UmuD heterodimers, which are more stable than either homodimer and all three dimers interact with UmuC. UmuD′ does not act in the DNA damage checkpoint control, but rather partici-

pates in the second role of the *umuCD* gene products. The UmuD′$_2$ homodimer interacts with UmuC, forming the UmuD′$_2$C complex (Pol V) that functions in a lesion-bypass DNA polymerase to enable *trans*lesion DNA *s*ynthesis (TLS; see Section 4.1.4) in the presence of RecA and SSB. While UmuC contains the catalytic DNA polymerase activity, the UmuD′$_2$ complex manages the activity of UmuC. Pol V has the remarkable ability to copy over abasic sites, cyclobutane dimers and pyrimidine-pyrimidone photoproducts. This ability, however, comes at the cost of reduced fidelity. Thus, replication by Pol V is inherently less accurate, leading to the formation of mutations, even when replicating undamaged templates, which is called SOS mutagenesis.

Post-translational modification of UmuD to yield UmuD′ acts as a molecular switch to regulate temporally the two physiological roles of the *umuCD* gene products. In addition, specific interactions of UmuD and UmuD′ with the α (catalytic), ε (proofreading) and β (processivity clamp) subunits of the *E. coli* replicative DNA polymerase, Pol III, are believed to be important for regulating which role the *umuCD* gene products will play. UmuD′$_2$C is also uniquely capable of inhibiting RecA-mediated recombination both *in vivo*, when UmuD′$_2$C is constitutively expressed at elevated levels, and *in vitro*. This may serve to turn off recombination in favor of translesion synthesis. Once SOS mutagenesis is no longer required, Pol V activity must be switched off quickly. UmuD′-UmuD heterodimers are key to this process: once regions of DNA that contain lesions are filled in, RecA::single-stranded DNA nucleoprotein filaments do not form, UmuD is not cleaved and UmuD′-UmuD heterodimers form that inhibit translesion synthesis. Heterodimer formation also targets UmuD′ for degradation by the ATP-dependent protease ClpXP, completely removing UmuD′ from the cell. Therefore, UmuD functions as a UmuD′ delivery protein for ClpXP (see also under 7.5.1). UmuD$_2$ is then degraded by Lon protease, shutting down the entire system.

Throughout this intricately regulated process, there are a myriad of critical protein-protein intercations. These include: (a) formation of UmuD$_2$ and UmuD′$_2$ and their complexes with UmuC, (b) the interaction of UmuD with RecA::single-stranded DNA filaments required for cleavage to UmuD′, (c) interactions of UmuD$_2$ and UmuD′$_2$ with Pol III subunits, (d) UmuD′$_2$C interactions with RecA, SSB and Pol III during translesion synthesis, (e) UmuD′$_2$C interactions with RecA that inhibit homologous recombination and (f) the interactions of UmuD$_2$ and UmuD′-UmuD with Lon and ClpXP. How do the structures of the two small proteins UmuD′ and UmuD (115 and 139 residues, respectively) enable them to participate in so many important interactions?

5.3.10
Replication and Repair

Recently, a family of error-prone DNA polymerases, named the Y family, was discovered. Found in all three domains of life, many of these polymerases can replicate past DNA lesions, but at the cost of frequent mutations. *E. coli* is known to have two Y-family DNA polymerases, Pol IV, the product of the *dinB* gene (also

called *dinP*), and Pol V, the product of the *umuCD* operon. Both of these polymerases are induced after DNA damage as part of the SOS response (see above). Whereas PolV can replicate past a variety of DNA lesions, the ability of Pol IV to do so is modest; and its bypass ability clearly depends on the lesion and the sequence context. To keep spontaneous mutation rates low, the level and activity of PolV are tightly controlled and targeted in *E. coli*. In contrast, the levels of Pol IV are relatively high. In the absence of DNA damage, normal cells have about 250 copies of Pol IV enzyme, whereas it is estimated that there are only 15 copies of Pol V in uninduced cells. Unlike Pol V, Pol IV has no cofactors. *In vitro*, its efficiency and processivity are dramatically increased by the β clamp; and it is possible that *in vivo* Pol IV synthesizes tracts of error-containing DNA up to 1300 bases long. Overproduction of Pol IV is a powerful mutator in growing and in stationary phase cells. However, loss of Pol IV has only modest effects on normal growth-dependent mutation rates. These results suggest that, in normal cells, there are mechanisms to keep the mutagenic activities of Pol IV under control.

The rescue of replication forks stalled by unpaired lesions in the template DNA is critical for efficient replication and transmission of the genome. It involves two essential events. First, the block to fork progression must be removed or bypassed. Second, replication has to be restarted. In *E. coli*, restart relies on the PriA protein to load the DnaB replicative helicase at branched DNA structures, which implies that stalled replisomes do not remain intact and have to be rebuilt. Two restart models have been proposed, the *direct rescue* and the *recombination model*. Both models are initiated by regression of the stalled fork to form a Holliday junction.

The direct rescue model requires either the RecG helicase or the RecBCD exonuclease to correct the replication fork. Next, the helicase function of PriA unwinds the lagging strand of the replication fork, facilitating the loading of DnaB on the lagging strand template via the primosome assembly function of PriA. The DnaG primase synthesizes the primer on the lagging strand, which is elongated by Pol III. The recombination model involves fork breakage, generation of single-stranded 3′-ended DNA through the action of RecBCD, loading of RecA and invasion of the second double-stranded DNA. The Holliday structure is resolved by by RuvABC, followed by loading of DnaB on the lagging strand through PriA (Fig. 5.9).

DNA lesions that have escaped repair can now block replication and thereby jeopardize genome integrity. When replication stops at lesions (estimated range from 10% to 50% of cells requiring some form of fork reactivation), single-stranded DNA regions carrying the damaged nucleotides are formed. In *E. coli*, these gap–lesion structures can be filled in with patches from the sister chromatids by recombinational repair, in an *error-free* manner. Alternatively, the gap can be fiilled by DNA synthesis called *trans*lesion *s*ynthesis (TLS), a process which is *error-prone* (mutagenic) due to the miscoding nature of most damaged nucleotide. Both mechanisms restore the double-stranded configuration of the DNA at the damaged site, thereby enabling a second attempt of excision repair.

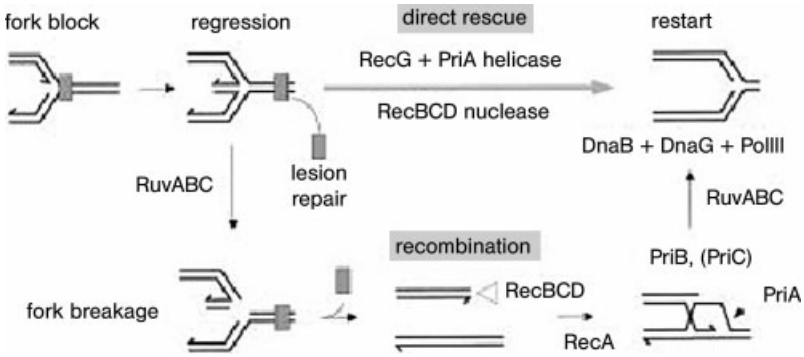


Fig. 5.9 Models for the rescue of replication forks stalled at a DNA lesion. The direct rescue model needs correction of the replication fork by either RecG or RecBCD followed by loading of the DnaB helicase on the lagging strand by PriA. The recombination model starts with fork breakage by the RuvABC complex, production of 3′-ended single-stranded DNA through the RecBCD activity, loading of RecA, invasion of the second double-strand DNA molecule, resolution of the Holliday junction through the RuvABC resolvasome and loading of the DnaB helicase on the lagging strand by PriA. R.P. Jaktaji **2003**, *Mich. Med.* 47, 1091–1100; Fig. 1.

Recombination Repair

In all of the DNA repair systems outlined above, the cell removes the DNA damage from the DNA, using the information of the complementary strand to restore the correct sequence. But what happens if the damage occurs on both strands of the DNA at the same place? Or if both strands of the DNA are broken simultaneously at the same site? Mechanisms that allow the cell to tolerate DNA damage without ever repairing it are called *damage tolerance* mechanisms.

5.3.11
Repair of Interstrand Crosslinks in DNA

Interstrand crosslinks are likely to be the crucial cytotoxic lesion, formed by a variety of classes of effective antitumor agents, including the mustards, mitomycin C, cisplatin, ethyl methansulfonate and light-activated psoralens. These interstrand crosslinks lead to the joining of two bases in the opposite strands of the DNA. Interstrand crosslinks are repaired by *n*ucleotide *e*xcision *r*epair (NER; as shown in Fig. 5.10). In the first step, the UvrABC complex nicks one strand on either side of the interstrand crosslink, followed by moving out the short single-stranded DNA. This leads to a gap opposite the DNA damage. In the second step, the gap is filled by the activity of Pol II. In the last step, the damaged strand can be repaired by the nucleotide excision pathway.

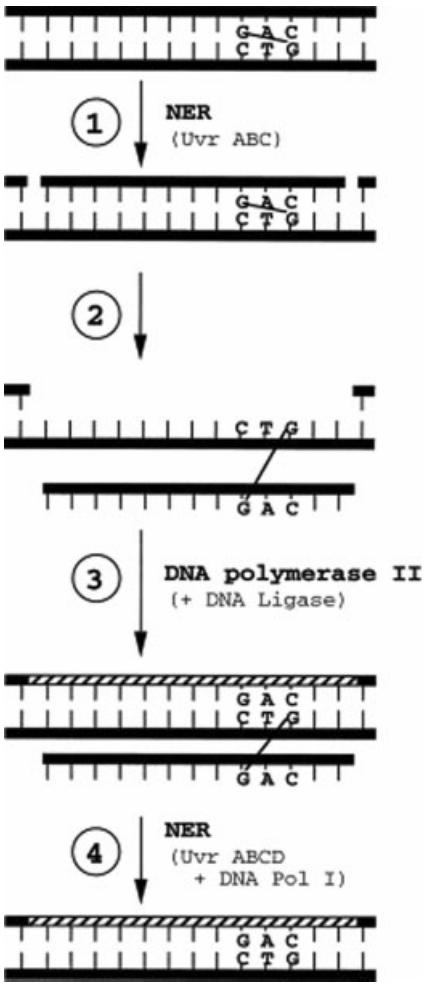


Fig. 5.10 The NER/DNA polymerase II pathway for the repair of interstrand crosslinks. The crosslink is recognized by the NER (UvrABC) complex which nicks one strand on both sites of the crosslink (1). The incised strand is moved out of the DNA (2), and the gap filled in by Pol II and sealed by DNA ligase (3). Finally, the monoadduct in the bottom strand is removed by NER (4). M. Berardini, et al. **1999**, *J. Bacteriol.* 181, 2878–2882; Fig. 1.

5.3.12
Deinococcus radiodurans

D. radiodurans is a mesophilic, nonspore-forming, nonmotile, spherical Gram-positive aerobic bacterium with a genome size of 3.2 Mbp that forms pairs and tetrads when grown in rich liquid medium. Cells divide alternately in two planes and optimal growth occurs at 30 °C. Cells grown in rich media are 'multigenomic', containing no less than four genome equivalents per cell (four to ten identical copies). This bacterium has an extraordinary resistance to γ-radiation and a wide range of other DNA-damaging conditions, including desiccation and oxidizing agents. Ionizing radiation induces DNA double-stranded breaks, which are the most lethal form of DNA damage. After acute exposure to 10 kGy, early stationary

phase cells can reassemble their genome from hundreds of DNA fragments without lethality or induced mutagenesis. *D. radiodurans* is even able to grow at 60 Gy h^{-1} without any discernable effect on its growth rate. Even after irradiation with 20 kGy, 3000 times the human deadly dose, surviving cells can be isolated. The ability of *D. radiodurans* to survive ionizing radiation raises an intriguing evolutionary question: why would this capability evolve in any species? Resistance to ionizing radiation does not provide a selective advantage, as there are no terrestrial environments that generate a radiation flux in excess of 200 mGr $year^{-1}$. The radiation resistance of *D. radiodurans* is a consequence of its ability to survive desiccation. Desiccation causes a time-dependent increase in DNA double-stranded breaks, which are the most deleterious lesions formed by ionizing radiation.

If *D. radiodurans* is strongly irradiated, its chromosomes will break in thousands of pieces. The only known mechanism that enables accurate repair of double-stranded breaks in bacteria is RecA-dependent homologous recombination, whereby information lost as a lesion is restored by a homologous DNA sequence that acts as a template (see Section 4.1). As such, DNA repair via homologous recombination strictly depends upon the ability of cellular systems to carry out a rapid and genomewide search for homologous DNA sites. However, in the present case, following extensive DNA fragmentation, no intact template remains. Homologous search performed under these circumstances would entail repetitive inspection of multiple randomly dispersed DNA fragments. Nevertheless, the cells will survive since they are able to put together all these pieces of DNA like a puzzle. How does *D. radiodurans* accurately reconstruct its genome from hundreds of radiation-generated fragments in the absence of an intact template? It has been shown that the *D. radiodurans* genome assumes an unusual toroidal (ring-like) morphology restricting diffusion within the tightly packed and laterally ordered DNA toroids. This unusual morphology suggests that radiation-generated free DNA ends are held together, which may facilitate template-independent error-free joining of DNA breaks.

It is believed that *D. radiodurans* has a DNA repair strategy that is quite different from that of other, better characterized prokaryotes. For its survival, *D. radiodurans*: (a) relies on homologous recombination, thus taking advantage of its multiple chromosomes, (b) uses post-irradiation DNA degradation to provide single-stranded DNA for recombination, (c) regulates DNA replication and post-irradiation DNA degradation to maximize repair efficiency and (d) might use the export of damaged nucleotides from the cell to avoid mutation. DNA replication, degradation and recombination repair appear to be coordinately regulated, suggesting that these processes are sensitive to intracellular signals.

DNA toroids are not specific for *D. radiodurans*, but have been described also for additional members of the family Deinococcaceae and were shown to persist in germinating spores of both *B. subtilis* and *B. megaterium*. The DNA-binding protein HU and *s*mall DNA-binding, *a*cid-soluble *p*roteins (SASPs) specifically stabilize toroidal DNA packing in *Deinococcus* and *Bacillus*, respectively.

6
Principles of Gene Regulation

Regulation of gene expression can occur at different levels. The first level is the DNA itself where particular conformations such as bends and looped DNA or superhelicity can influence the expression of some genes. Another important mechanism is DNA modification, including the methylation of particular bases such as the adenine residues in the sequence GATC by the Dam methylase. Yet another form of gene regulation occurs by decryptification of so-called silent or cryptic genes. The second and most important level of regulation occurs at the level of transcription, including initiation, elongation and termination. Epistatic on these mechanisms are a plethora of anti-termination and attenuation mechanisms. Initiation of transcription is often regulated by transcription factors which act either as repressors or as activators (alternative sigma factors, or true transcriptional activators). Post-transcriptional mechanisms involve stability of transcripts, genetically regulated processing and interaction with regulatory RNAs, to mention some. A third level of gene regulation can occur at the level of translation. And the fourth level involves the proteins themselves, including their half-life and their interaction with cofactors, including other proteins.

6.1
Regulation at the Level of DNA

Regulation of gene expression at the level of DNA is rare. It involves changes in the superhelicity of the DNA (studied extensively using plasmid DNA), intrinsic and protein-induced bents upstream of promoters and often results in DNA looping if two binding sites separated by a certain distance occur. More important are, at least in *E. coli*, DNA methylation at GATC sites, often within the promoter region, and activation of silent genes by a mutational event.

Dynamics of the Bacterial Chromosome. Structure and Function. W. Schumann
Copyright © 2006 Wiley-VCH Verlag GmbH & Co. KGaA
ISBN 3-527-30496-7

6.1.1 Alterations of the Structure of the DNA

Superhelical Density-dependent Promoters

The structure of purified DNA is sensitive to environmental changes. It has been shown that the helical pitch (number of base pairs per turn) decreases when the temperature increases and increases when the concentration of certain salts increases. Changes in the pitch effect DNA supercoiling, which is normally corrected by topoisomerases trying to maintain the intracellular tension at a correct value. But some environmental factors change supercoiling. Entry into stationary phase and growth at low pH reduce supercoiling, while growth at high osmolarity or low oxygen tension increase supercoiling. How can these factors influence supercoiling? It has been observed that there is a correlation between the extent of supercoiling and the ratio of ATP to ADP. While a reduce in supercoiling occurs after a drop of the ATP:ADP ratio, the contrary occurs after an increase. Does the level of supercoiling influence transcription? While some genes are activated, others are inhibited and others still are unaffected by the same supercoiling change. Underlying such complexity is the fact that supercoiling can affect the DNA helix in various ways, by modifying its energy (torsional strain) and structure (helical pitch and axial writhing). Any of these effects can influence promoter activity either directly or indirectly through effects on the bending and wrapping of the DNA around proteins in chromatin-like structures. One might expect that, at least in some cases, the information specifying a given response is encoded in the promoter primary sequence. One studied example is the *gyrA* promoter of *E. coli*, whose activity is stimulated by gyrase inhibition causing DNA relaxation. The minimal sequence for the response lies within a 20-bp segment spanning positions –19 to +1 of the promoter region.

DNA Curvature (Bending)

A naturally occurring bent DNA segment was discovered in kinetoplast DNA minicircles from the parasite *Leishmania tarentolae* in 1982 due to its unusually slow electrophoretic mobility in nondenaturing polyacrylamide gels. Since then, bent DNA structures have been identified in a wide variety of cellular and viral genomes from bacteria to man. Today, we distinguish between *intrinsic bends* and *protein-induced bends*. Intrinsic bends form when special base sequences, usually short (3–6 bp) runs of adenines, called A-tracts, occur repeatedly in phase with the DNA helical repeat of about 10.5 bp. Two different kind of models have been presented to explain the molecular basis of DNA bending: *junction models* and *wedge models*. Junction models assume an abrupt change in the DNA helix axis direction occurring at the junction of two distinctly different DNA helices, whose base pairs have different angular orientations relative to their respective helix axes. Wedge models suggest smooth global bending as the result of the vectorial sum of all dinucleotide wedges, which are composed of the roll and tilt components along the length of the molecule.

Bent DNA appears to be implicated in both activation and repression of transcription. In activation, it seems to play a direct role while, in repression, a *silencer protein* that binds to it plays a crucial role. In transcription activation, bent DNA often occurs in the region from about –100 to about –35, relative to the transcription initiation site. Such an intrinsic bend facilitates the binding of RNA polymerase to promoters and/or the formation of an open promoter complex. Bent DNA immediately upstream of a promoter usually has a right-handed superhelical form, but the wrapping of DNA on RNA polymerase is inferred to be left-handed. Therefore, the binding of RNA polymerase would change the sense of the superhelical twist from right- to left-handed, generating a structural strain leading to deformation of double-stranded DNA. This in turn can cause local unwinding, which might facilitate transcription. The positive effect of a right-handed twist was verified for the promoters upstream of the β-lactmase gene (*bla*) in pUC19 and the tetracycline resistance gene (*tetR*) of pBR322. In transcription repression, DNA bends play an indirect role by acting as a binding site for specific silencer proteins. They stabilize or enhance a preexisting DNA loop, thus effectively blocking transcription of downstream regions. H-NS is the most studied silencer protein, affecting a large number of unrelated enterobacterial genes coding for housekeeping functions as well as for virulence factors. The molecular basis for the H-NS regulatory activity probably resides in its preferential interaction with intrinsically curved DNA and its ability to induce the bending of noncurved DNA. The well studied example of H-NS silencing the *bgl* operon is described below. Another example is the *virF* promoter of *S. flexneri*. This promoter controls transcription of the *virF* gene coding for a transcriptional activator of the virulence genes. Since expression of the virulence genes is only required after infection of a mammalian host such as humans, but not in habitats such as water, how does *S. flexneri* know when it has entered a mammalian host? The pathogenic bacterium senses the temperature (37 °C in a human host, below 30 °C in other habitats). Sensing of the temperature occurs via a DNA bend and H-NS protein binding to both ends of the bend. It has been shown experimentally that the curvature is directly dependent on the temperature. While there is a sharp bent at 4 °C, the DNA molecule is almost straight at 60 °C. This was measured in the absence of H-NS, which additionally influences the extent of the bend.

DNA Looping

DNA looping is defined as an extreme bend in DNA that allows two distinct regions to come into close proximity. Proteins bound at sites separated by a distance can then directly interact to effect transcription. Very often, the spacing of distant protein binding sites by integral helical turns is important. The addition or deletion of half a turn of the helix places the two sites out of phase, significantly influencing transcription. Two well studied examples of enhancing transcription repression are the *lac* and *ara* operons of *E. coli*.

The *lac* operon region contains a total of three *lac* operators, the primary operator (O_1) overlapping with the promoter (Fig. 6.1) and two secondary operators

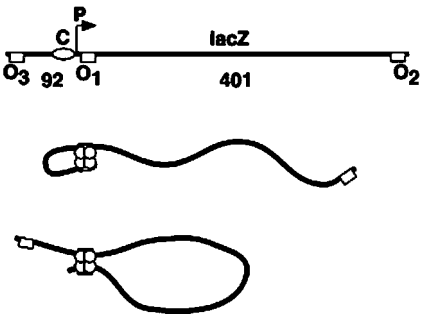


Fig. 6.1 The *lac* operators of *E. coli* and deduced DNA loops. The *lac* operon region contains three *lac* operators, the major operator O_1 located immediately downstream of the *lac* promoter and two minor operators, where O_2 is located within the coding region of *lacZ* and O_3 within that of *lacI*. The tetrameric LacI repressor first interacts with O_1, and then with either O_2 or O_3 to form looped DNA thereby enhancing repression. N.A. Becker, et al. **2005**, *J. Mol. Biol.* 349, 716–730; Fig. 1a, b.

(O_2, O_3). O_2 is located within the coding region of *lacZ*, 401 bp downstream of O_1, while O_3 is situated within the coding region of *lacI*, 92 bp upstream of O_1. *In vitro* studies have shown that the LacI repressor can occupy two operators simultaneously, forming a loop as shown in Fig. 6.1. The LacI repressor first binds to O_1 as a tetramer and then contacts either O_2 or O_3. While binding to O_1 alone results in a 20-fold repression, binding to two operators simultaneously increases repression to 50-fold.

The AraC protein possesses two functional domains: a dimerization domain also involved in binding L-arabinose located within the N-terminal part and a DNA-binding domain situated near its C-terminus. Both domains are loosely connected by a flexible linker, allowing the active homodimer to interact with its binding sites. In the absence of L-arabinose, AraC binds to two sites termed O_2 and I_1, forming a DNA loop as shown in Fig. 6.2.

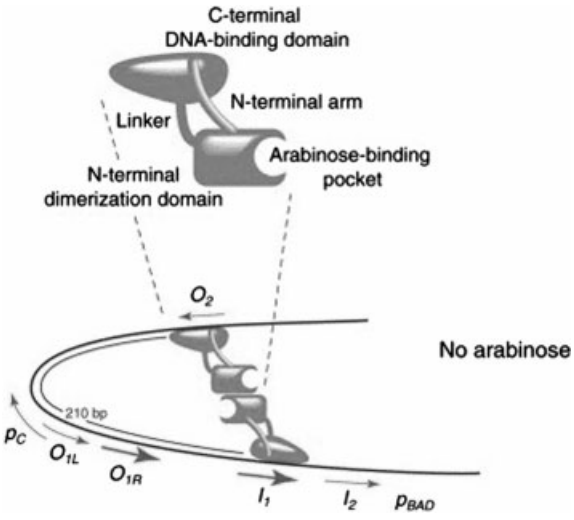


Fig. 6.2 Domain structure of the AraC transcriptional regulator and loop formation in the absence of arabinose. The AraC protein consists of two functional domains, an N-terminal dimerization domain containing an arabinose-binding pocket and a C-terminal DNA-binding domain; both domains are connected by a flexible linker. In the absence of arabinose, the AraC dimer binds to the sites O_2 and I_1 forming a loop resulting in repression of the P*BAD* promoter. R. Schleif **2000**, *Trends Genet.* 16, 559–565; Fig. 2.

6.1.2 GATC Methylation and Gene Expression

While specific methylation at the C5 position of cytidines plays an important role in the regulation of eukaryotic genes, it plays only a minor role in eubacteria. In *E. coli*, there are two genes coding for DNA methyltransferases, *dam* (*D*NA *a*denine *m*ethylation) and *dcm* (*D*NA *c*ytosine *m*ethylation). While the Dam methylase recognizes the sequence GATC and methylates the adenine residue, Dcm recognizes CCA/TGG and methylates the second cytosine residue. Immediately after the replication fork has passed a GATC sequence, it is present in a hemimethylated way which plays an important role during the regulation of initiation of replication (see Section 3.1.1.6) and methyl-mediated mismatch repair (see Section 5.3.1). In addition, GATC methylation can influence gene expression by two different mechanisms. First, methylated GTAC sequences within the –10 or the –35 element of the promoter can increase (e.g., *dnaA*), decrease (e.g., *sulA*) or have no effect on transcription, or the hemimethylated status can enhance transcription (see Table 6.1). Second, the Dam methylase and regulatory proteins, such as CRP, Lrp or OxyR, compete for overlapping sites in or near promoters. Here, regulation of the *pap* operon is the best studied example.

Let us first consider the four examples presented in in Table 6.1. Methylation sites are present in the –35, the –10 or both regions. The P_2 promoter is present in front of the *dnaA* gene coding for the replication initiator protein, *glnS* codes for the glutaminyl tRNA synthetase, *trpR* for the tryptophan repressor and *tnpP* for the transposase of the mobile elements Tn*5*. What could be the biological function of coupling gene regulation to the methylation of GATC sequences? As mentioned, these sequences occur in a hemimethylated state for some minutes when the replication fork has passed over. Under these conditions, the RNA polymerase binds more efficiently to these promoters, thereby coupling gene expression to the cell cycle. This becomes evident for *dnaA*, which needs increased expression once during the cell cycle, but is less evident for *glnS* and *trpR*. In the case of the transposase gene, hemimethylated promoter sequences are more active than fully methylated sequences; and completely unmethlyated GATC sites further increase the promoter strength. The latter case often occurs when a transposon is introduced into *E. coli* cells from another species not encoding the Dam methylase. If, e.g., the transposon is part of a transducing DNA which fails to re-

Table 6.1 Dam-methylation sensitive promoters. The GATC sequences are given in italics.

Gene	*Promoter (from –35 to –10)*
dnaA P_2	AGAA*GA TC*TCTTGCGCAGTTTAGGC TAT*GAT C*
glnS	TTGTCA GCCTGTCCCGCTTATAA *GATC*AT
trpR	CT*GATC* CGCACGTTTATGATATGC TATCGT
tnpP Tn*5*	GGAACC TTTCCCGTTTTCCA*GA TC*T*GAT C*

combine with the chromosome, it can be rescued by transposing from the foreign DNA into the *E. coli* chromosome.

The *pap* (*p*yelonephritis-*a*ssociated *p*ili) operon is present in *u*ro*p*athogenic *E. coli* (UPEC) strains. The *pap* promoter controls the expression of the Pap in UPEC cells, allowing UPEC to bind to uroepithelial cells. The *pap* regulatory region encompasses the divergently transcribed genes *papI* and the *papB* operon, together with a 416-bp intergenic region (Fig. 6.3). While the 8-kDa PapI protein acts as a coregulatory protein together with Lrp (*l*eucine *r*esponsive *p*rotein), the *papB* gene codes for another 12-kDa regulatory protein. The P pilus is composed of several subunits and its assembly is described under 7.2.2. These pili play an essential role in colonization of the upper urinary tract. There are two GATC sequences and six binding sites for Lrp located within the intergenic region, where sites 1–3 are located at and around the *papB* proximal GATC site and sites 4–6 are at and around the distal site (Fig. 6.3). When the GATC sequence proximal to *papB* is occupied by Lrp, transcription of *papB* is turned off. When the distal site is bound by Lrp together with the PapI corepressor, *papB* is in the ON state. Which site is occupied is dictated by the methylation status of the GATC sequence, since Lrp or Lrp-PapI bind only to nonmethylated sites. The switching between the ON and OFF states constitutes an *epigenetic switch*. In the OFF state, the proximal GATC sequence is unmethylated and the distal sequence is fully methylated. Therefore, Lrp binds to the proximal sequence at sites 1–3, where site 3 overlaps with the promoter directing expression of the *papB* operon. When the replication fork has moved over this region, both GATC sequences are present in a hemimethylated state. This can cause a shift of Lrp to sites 4–6, which is exaggerated by binding PapI. This ON state is further stabilized by full methylation of the proximal GATC sequence. After another round of replication, again an unstable situation is created where Lrp may bind again predominantly to sites 1–3, switching this cell into the OFF state. A factor that may contribute to the OFF state is a mutual exclusion phenomenon. The affinity of Lrp is about ten times higher for sites 1–3 compared with sites 4–6, indicating that binding of Lrp at sites 1–3 exerts a negative effect on Lrp binding at sites 4–6. In contrast, PapB helps maintain the

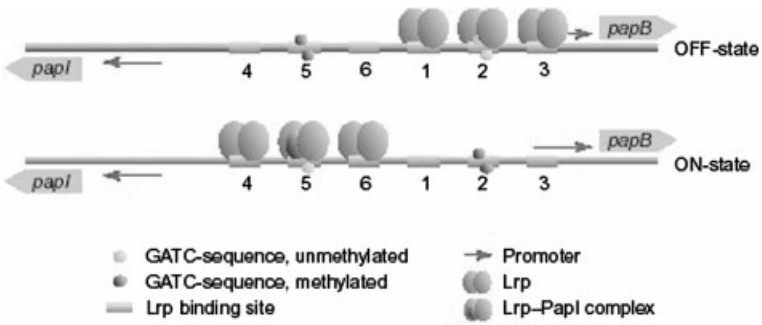


Fig. 6.3 Turning *pap* pilus synthesis on and off. See text for explanation how a cell can shift from the OFF- to the ON-state and vice versa. A. Løbner-Olesen, et al. **2005**, *Curr. Opin. Microbiol.* 8, 154–160. (This figure also appears with the color plates.)

switch in the ON state by activating *papI* transcription, which in turn maintains Lrp binding at sites 4–6. In the OFF state, cells are nonpiliated, while in the ON state they are covered with pili. What are the biological consequences of this epigenetic OFF/ON switch? This complex mechanism allows for global (Lrp, Dam) as well as local (PapI, PapB) regulatory inputs to influence the Pap phase switch. In one specific environment, it might be an advantage to express Pap pili, while in another the OFF state might be more appropriate.

A. Hernday, et al. **2002**, Self-perpetuating epigenetic pili switches in bacteria, *Proc. Natl. Acad. Sci. USA* 99, 16470–16476.
A. Løbner-Olesen, et al. **2005**, Dam methylation: coordinating cellular processes, *Curr. Opin. Microbiol.* 8, 154–160.

6.1.3
Programmed DNA Rearrangements Within the Chromosome

DNA arrangements include duplications and amplifications, inversions and deletions. If these arrangements are genetically programmed, they can influence gene expression as already reported for amplifications (Section 2.4.1), deletions (Section 2.4.2) and inversions (Section 4.2.3).

6.1.4
Transcriptional Silencing and Cryptic Genes

Transcriptional silencing is a process rendering an appreciable region of DNA inaccessible to RNA polymerase. It involves one or more DNA *silencer* sequences to which a protein will bind, mostly the abundant, nucleoid-associated protein H-NS. Well studied examples of silencing are the *bgl* operon of *E. coli* and centromer-associated bacterial silencing. *Cryptic genes* are inactivated by a point mutation resulting in the synthesis of inactive proteins. Here, the *ebg* and the *ilvGM* operons have been analyzed in detail. While silenced genes are not expressed at a detectable level, cryptic genes are expressed. To obtain expression in the first case and synthesis of a functional protein in the second, a mutation is needed in both cases. Silenced genes are normally activated by insertion of a mobile element into the binding site, preventing binding of the protein to the silencer, while cryptic genes are decryptified normally by a point mutation, leading to the synthesis of an active protein.

The *bgl* Operon

Both laboratory strains and natural isolates of *E. coli* posssesses silent genes for the utilization of β-glucoside sugars. These genes are not expressed at a detectable level in wild-type organisms and must be activated by mutations (which may include insertion of IS or other mobile elements) before they can be expressed. Because silent genes are not expected to make any positive contribution to the fitness

of the organism, one would imagine that they are lost because of the accumulation of inactivating mutations. It follows that silent genes should be rare in natural populations. This is not the case, since >90% of natural isolates of *E. coli* carry silent genes for the utilization of β-glucoside sugars. The current model is that they are retained by alternately selecting for loss and regain of function in different environments. The paradigm example of silent genes in *E. coli* is the *bgl* operon involved in the uptake and degradation of aryl-β,D-glucosides, such as salicin, arbutin and cellubiose. The gene products of the *bgl* operon are the positive regulator and anti-terminator BglG, the β-glucoside-specific permease EII^{Bgl} (or BglF), and the phospho-β,D-glucosidase BglB. A key element in the silencing of the *bgl* operon is the global repressor protein H-NS, which represses the *bgl* operon ~100-fold. Repression requires an AT-rich and presumably bent silencer sequence located directly upstream of the CRP-dependent *bgl* promoter and a downstream silencer located within the coding region of the first gene, *bglG* (Fig. 6.4). This region located 600–700 bp downstream of the promoter, which is predicted to be bent, too. H-NS prevents expression of the *bgl* operon by repression transcription initiation at the promoter and by repressing transcription elongation. Here, H-NS acts as a roadblock to the elongating RNA polymerase, allowing termination factor Rho to catch up with the paused RNA polymerase and to terminate transcription. In addition, H-NS can trap two DNA strands together, with the intervening DNA forming a loop.

No laboratory conditions are known which relieve silencing. Silencing is overcome by spontaneous mutations mapping close to the CRP-dependent promoter, including the deletion of an AT-rich silencer upstream of the promoter, integration of IS elements and point mutations that impose the CRP-binding site. If the promoter-upstream H-NS binding site is destroyed by a mutation (mostly by integration of an IS element), RNA polymerase can bind and initiate transcription. In the absence of a β-glucoside, transcription is terminated at sites flanking the first gene of the *bgl* operon, *bglG* (see Section 6.2.7). If a β-glucoside is present, transcription will be elongated, resulting in the dissociation of the H-NS complex bound within *bglG*.

E. coli strains carry another cryptic operon involved in β-glucoside degradation, the *cel* operon. This operon consists of five genes: *celA–celF*. *celA*, *celB* and *celC* en-

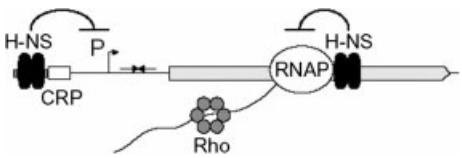


Fig. 6.4 Model of the H-NS-mediated silencing of the *bgl* operon. The H-NS protein binds to two AT-rich and most probably bent regions, one located upstream of the CRP-dependent promoter and the other within the coding region of the *bglG* gene. While through binding to the promoter region it represses initiation of transcription, interaction with the second site results in repression of transcription elongation. Here, H-NS acts as a roadblock allowing the Rho factor to catch up with the paused RNA polymerase resulting in its dissociation. S Dole, et al. **2004**, *Mol. Microbiol.* 52, 589; Fig. 7.

code, respectively, the PTS enzyme IV_{cel}, enzyme II_{cel} and enzyme III_{cel}, for transport and phosphorylation of cellobiose, arbutin and salicin. Gene *celD* encodes the Cel repressor; and *celF* encodes a phospho-β-glucosidase that hydrolyzes phosphorylated cellobiose, arbutin and salicin. The *cel* operon can be activated by insertion of IS*1*, IS*2* or IS*5* into a 108-bp region 72–180 bp upstream from the transcription start site. In addition, it can be activated by base substitutions in *celD* that alter the repressor so that it can recognize cellobiose, arbutin and salicin as inducers.

Silencing of Centromer-associated Regions

Large plasmids such as the F factor encode so-called partition modules that ensure their stable maintenance despite their low copy number (one plasmid copy per chromosome). The F factor partition module consists of the autoregulated *sopABC* operon, where *sopA* codes for an ATPase and *sopB* for a DNA-binding protein with affinity for sites involved in partitioning, termed *sopC*. This set of DNA sites (12 altogether) is also designated the plasmid centromere. The SopB protein not only binds to the *sopC* sites, but can also spread into the adjacent DNA sequences, causing silencing of genes flanking the centromere as far as 10 kb away. *In vivo* cross-linking experiments demonstrated that SopB contacted DNA along several kilobases on either side of *sopC*. What might be the biological function of this silencing? Genes susceptible to silencing are capable of being transcribed only when the replication fork has swept SopB away. Therefore, these genes can be transcribed only during a very short time during the cell cycle.

The Cryptic *ebg* Operon

In 1973, J. Campbell and coworkers published data showing that a Δ*lacZ* mutation could revert to Lac^+. They streaked a Δ*lacZ* strain onto lactose fermentation indicator plates supporting the growth of Lac^- cells and upon which Lac^+ and Lac^- cells formed differently colored colonies. After several weeks of incubation, Lac^+ papillae grew from the surface of many of the colonies. Upon restreaking onto lactose minimal medium, the formed slowly growing colonies exhibiting faster-growing papillae. These mutants were called *ebg* (for *e*volved *b*eta-galactosidase) and were mapped far away from the *lac* operon. Later, it turned out that the *ebg* region is organized into two transcriptional units, where one codes for a transcriptional repressor (327 amino acids; *ebgR*) and the second, located immediately downstream, for the new β-galactosidase (1030 amino acids; *ebgA*) followed by a gene of unknown function (*ebgC*). Sequencing of the mutants revealed that two mutations are needed for the synthesis of a functional β-galactosidase. One mutation has to inactivate the repressor so that it does not bind lactose, allolactose or the gratious inducer IPTG (but IPTG is used to induce the *lacY* gene allowing uptake of lactose), to allow constitutive expression of *ebgA*. And the second mutation has to alter the catalytic properties of the β-galactosidase. Depending on the location of the point mutation, the mutant enzymes allow growth either on lactose only or on lactose and lactulose. These enzymes could be further evolved to allow

growth on another lactose derivative, galactosyl-β-D-arabinose, a xenobiotic sugar not found in nature.

The Cryptic *ilvGM* Operon

Five isoenzymes catalyze the first step of the parallel valine–isoleucine biosynthetic pathway in *E. coli*. These enzymes are acetohydroxy acid synthase (AHAS) I to V (Table 6.2). Three enzymes are composed of a large (60 kDa) and a small (10–17 kDa) subunit (AHAS I, II, III). Furthermore, two enzymes are subject to feedback inhibition by valine, while the other three are not. Most importantly, three transcription units coding for AHAS enzymes are cryptic. In the case of *ilvG*, this gene contains a frameshift mutation, resulting in an inactive gene product of 300 amino acid residues. It has been shown that this gene can be decryptified by the addition of two base pairs, extending the gene product to 521 amino acids, being the active AHAS II enzyme.

B.G. Hall **2003**, Transposable elements as activators of cryptic genes in *E. coli*, *Genetica* 107, 181–187.

M. Yarmolinsky **2000**, Transcriptional silencing in bacteria, *Curr. Opin. Microbiol.* 3, 138–143.

Table 6.2 The AHAS enzymes of *E. coli*.

Gene	*Characteristics*	*Enzyme designation*
ilvBN	Expressed constitutively; valine-sensitive	AHAS I
ilvGM	Cryptic; valine-resistant	AHAS II
ilvIH	Expressed constitutively; valine-sensitive	AHAS III
ilvJ	Cryptic; valine-resistant	AHAS IV
ilvF	Cryptic; valine-resistant	AHAS V

6.2 Regulation at the Level of Transcription

Transcription is the first step in gene expression, it is the step at which most regulation of gene expression occurs and it is facilitated by a complex of proteins called RNA polymerase. RNA polymerase has a sequence-specific affinity for the region 5′ to the coding sequence, known as the promoter.

6.2.1 The DNA-dependent RNA Polymerase

RNA synthesis in all cellular organisms is mediated by the DNA-dependent RNA polymerase whose structure and function are conserved from bacteria to human. Two different forms of the bacterial RNA polymerase are distinguished, the *core* and the *holoenzyme*. The core enzyme has a molecular mass of ~400 kDa and a

conserved subunit composition of $\alpha_2\beta\beta'\omega$. It possesses catalytic activity which includes the synthesis of RNA complementary to the DNA template in the presence of nucleoside triphosphates and the degradation of nascent RNA through pyrophosphorolysis or hydrolysis. The core RNA polymerase by itself is incapable of recognizing specific promoter sequences, or of melting DNA and initiating transcription. To carry out these functions, the enzyme must bind a specificity factor, called σ, to form the holoenzyme. Bacteria code for two different kinds of σ factors, termed *primary* or *housekeeping* σ factor and *secondary* or *alternative* σ factor. There are about 2000 molecules of RNA polymerase per cell.

Function of the Different Subunits of the RNA Polymerase

The *α subunit* encoded by gene *rpoA*, consisting of 329 amino acid residues (36.5 kDa), is composed of two independently folded domains. The N-terminal domain (αNTD; residues 8–233) is responsible for dimerization and interactions with β and β′ (α^I interacts with β, α^{II} with β′) and the C-terminal domain (αCTD, residues 245–329), tethered to αNTD through a flexible linker consisting of at least 13 amino acid residues, contains DNA-binding determinants. Genetic analyses have identified amino acid residues in the αCTD critical for DNA binding and *UP element* function (see below). These residues are essential for cell viability and are highly conserved in prokaryotic α sequences. It has been suggested that αCTD is member of the helix-hairpin-helix (HhH) domain family of DNA-binding proteins. Furthermore, αCTD is the interaction target of a number of transcriptional factors, including cyclic AMP receptor protein (CRP), OmpR and GalR.

The *β subunit* encoded by the gene *rpoB* is 1342 amino acids in length (150.6 kDa) and, together with the β′ subunit, constitutes the catalytic core of the enzyme. This subunit is the target of the two important antibiotics, rifampycin and streptolydigin, which inhibit initiation and elongation of transcription, respectively. Rifampycin blocks synthesis of RNAs longer than two to three nucleotides. Furthermore, ppGpp binds to the β-subunit and affects both transcription initiation and polymerase pausing (see Section 9.7).

The *β′* subunit (*rpoC*; 1407 amino acids; 155 kDa) is the largest subunit and involved in catalysis.

The *ω subunit* (*rpoZ*; 91 amino acids; 10 kDa) is an often ignored polypeptide present in highly purified preparations of *E. coli* RNA polymerase. It is a conserved polypeptide which is present in all sequenced genomes of free-living bacteria, suggesting an important, conserved function. It has been suggested that ω functions in RNA polymerase assembly by "latching" the N- and C-terminal regions of the β′ subunit, thereby facilitating association of the RNA polymerase largest subunit with the $\beta\alpha^I\alpha^{II}$ assembly intermediate. This function can also be carried out by the GroEL chaperonin which serves as a backup system. Recent data suggest that the ω subunit also fulfils a function during the stringent response (see Section 9.7). During *in vitro* transcription, RNA polymerase without the ω subunit loses its responsiveness towards ppGpp. Under *in vivo* conditions, ω can be replaced by the DskA protein.

The *σ subunit* has three major functions: (a) It ensures the recognition of specific promoter sequences, (b) it positions the RNA polymerase holoenzyme at a target promoter and (c) it facilitates unwinding of the DNA duplex near the transcription start site. As already mentioned, with a few exceptions, all bacteria code for one *primary* or *housekeeping* sigma factor which is involved in the transcription of all those genes expressed during the exponential growth phase. The products of these genes are involved in DNA replication, transcription, translation, cell division, etc. In addition, almost all bacterial species contain one or more additional sigma factors, termed *secondary* or *alternative* sigma factors. These sigma factors are needed only under specific conditions, such as a sudden increase in temperature (see Section 9.2). While *M. genitalium* codes for only just one sigma factor, *E. coli* codes for a total of seven sigma factors (one housekeeping, six alternatives; see Table 6.3), *B. subtilis* for 17 (one housekeeping, 16 alternatives) and *S. coelicolor* even for 34. Alignment of all known sigma factors classifies them into two different families, the σ^{70} and the σ^{54} family, with little if any sequence identity between them. While almost all of them belong to the σ^{70} class, most, but not all, bacteria contain one class σ^{54} factor.

The structure of the *E. coli* RNA polymerase, obtained by electron crystallography at ~25 Å resolution, reveals an overall 3-D size of the enzyme of approximately 150×100×100 Å and the presence of three major channels:

- The primary channel, 25 Å in diameter and 55 Å in length that is proposed to comprise the DNA-binding site. Experiments have shown that a DNA length of ~70–95 bp is protected by the RNA polymerase.
- The RNA exit channel, which includes the single-stranded RNA.
- The secondary channel, which allows substrate diffusion, binds the extruded RNA 3′-terminus in backtracked ternary elongation complexes, connects the active site to the surrounding solution and may serve as a gate for entering nucleoside triphosphates, exiting pyrophosphates, or both.

Table 6.3 The σ factors of *E. coli*.

σ factor designation	*Function*
σ^{70}	Housekeeping σ factor
σ^{32}	σ factor activated by damaged cytoplasmic proteins
σ^{E}	σ factor activated by damaged periplasmic proteins
σ^{54}	σ factor activated under conditions of N starvation
σ^{FecI}	Required for synthesis of the ferric citric transporter
σ^{28}	Required for synthesis of flagella

The Promoter of the σ^{70} Type

σ^{70}-Dependent promoters contain two principal elements, the –10 (consensus sequence: TATAAT) and the –35 hexamers (consensus sequence: TTGACA), respectively, located about 10 bp and 35 bp upstream from the transcription start site

(defined as +1). Both hexamers are separated by a 17±1 bp spacer region with an optimal length of 17 bp. The two other important promoter elements are the *extended –10 region* and the *UP element* (see below).

The σ^{70} Family

The majority of σ factors belong to the σ^{70} family, named after the housekeeping σ^{70} of *E. coli*. Sequence comparison of σ^{70} family proteins led to the identification of four highly conserved amino acid regions, which have been further subdivided (Fig. 6.5). Subregion 1.1 occurs only in housekeeping σ factors; and this region prevents interaction of σ^{70} with DNA. Upon interaction with the core enzyme, a conformational change alters this region, allowing the holoenzyme to bind to its promoter. Region 1.2 seems to be required for open complex formation. Region 2.1 is important for interaction with the subunits of the core RNA polymerase, while the function of region 2.2 is unknown. Region 2.3 is involved in DNA melting, region 2.4 contains a HTH motif and recognizes the –10 promoter element and region 2.5 the so-called extended –10 region. Region 3.2 is also involved in recognition of the core enzyme, while region 4.1 can make contact with several transcriptional activators and region 4.2 contains the second HTH motif which recognizes the –35 region.

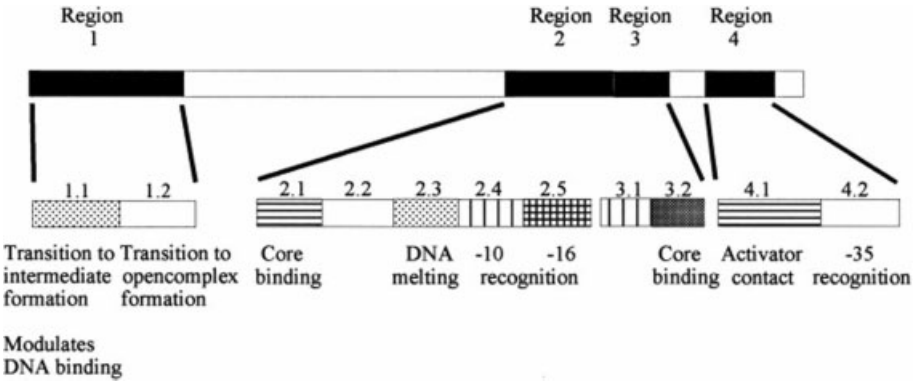


Fig. 6.5 The elements of σ^{70}. Sequence comparison among σ factors of the σ^{70} family revealed four major regions of sequence similarity which have been further subdivided. See text for detailed explanation of the different subregions. M.M. Wösten **1998**, *FEMS Microbiol. Rev.* 22, 127–150; Fig. 2.

Promoters of the σ^{54} Family

In contrast, members of the σ^{54} (also called σ^{N}) family differ in three aspects from those of the σ^{70} family: (a) their domain organization, (b) they recognize the –24 (GG) and –12 (GC) region of their cognate promoters and (c) they are completely unable to initiate transcription and are absolutely dependent on an activator protein for this first step of transcription, formation of the open complex. The σ^{54} fac-

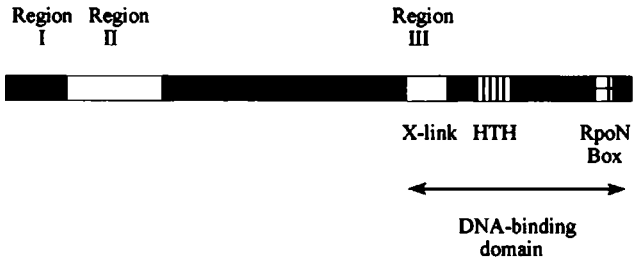


Fig. 6.6 The elements of σ^{54}. These sigma factors consist of the three regions I, II and III. See text for a detailed description of these regions. M.M. Wösten **1998**, *FEMS Microbiol. Rev.* 22, 127–150; Fig. 2.

tors consists of three functional domains, I, II and III (Fig. 6.6). Region I is glutamine-rich, comprises 25–50 amino acids and is not conserved. It has a regulatory function and is closely implicated in RNA polymerase isomerization and DNA melting. It plays a central role in mediating the response to activator proteins and in binding early melted DNA structures, but is dispensable for core and overall DNA binding. Region II contains 60–110 amino acids with a significant excess of acidic residues. Region III has a length of about 400 amino acids and consists of the X-link region (can be cross-linked to DNA) and two well conserved motifs, a HTH motif and the RpoN box, where both are involved in recognition of the promoter. The RpoN box is characterized by a stretch of ten conserved amino acids: ARRTVAKYRE. Region III includes a major core RNA polymerase-binding determinant of 95 amino acid residues and sequences that directly contact DNA and enhance σ^{54}-DNA binding. Emerging roles for region III appear to be in maintaining the closed promoter complex in a transcriptionally silent state and in generating polymerase isomerization upon activation.

The key elements of the σ^{54}-dependent promoter are located near positions –12 (GC) and –24 (GG). The σ^{54} holoenzyme binds to the promoter and forms a closed promoter complex, but is unable to proceed to open complex formation. Open complex formation is completely dependent on a transcriptional activator protein. These specialized activator proteins function as higher-order oligomers from *enhancer*-like sequences typically located 70–150 bp upstream of the transcription start site and contact the $E\sigma^{54}$ closed complex via a DNA looping event, which is often induced by binding a protein such as the IHF. The activators carry specific conserved modules that drive ATP-dependent remodeling of the RNA polymerase–promoter complex, resulting in the formation of the open complex. Well studied activators are NtrC in *Rhodobacter* spp and *K. pneumoniae* and NifA in *Rhizobium* spp, both involved in nitrogen fixation. FhlA activates the formate regulon in *E. coli* and XylR stimulates the degradation of aromatic compounds in *P. putida*.

Alternative Sigma Factors

The alternative sigma factors (which include σ^{54}) are needed only under specific conditions. This means that these factors are either synthesized only on demand or, if produced constitutiveley, held in an inactive state. The six alternative sigma factors of *E. coli* are involved in transcription of the following group of genes: $E\sigma^{54}$ (or $E\sigma^{N}$) transcribes genes which are regulated by the availability of nitrogen, $E\sigma^{32}$ the genes for heat-shock proteins (see Section 9.2), $E\sigma^{F}$ the genes needed for expression of flagella and chemotaxis, $E\sigma^{S}$ the genes active during stationary phase growth and the two ECF σ factors $E\sigma^{E}$ and $E\sigma^{FecI}$ the genes for *e*xtra*c*ytoplasmic *f*unctions, as well as the heat-shock response and iron uptake, respectively. All *E. coli* sigma factors and the promoters recognized by them are presented in Table 6.3. Mixed *in vitro* reconstitution experiments in the presence of a fixed amount of core enzyme and increasing amounts of an equimolar mixture of all seven σ subunits indicated that σ^{70} is strongest in terms of core enzyme binding, followed by σ^{N}, σ^{F}, σ^{E}/σ^{FecI}, σ^{H} and σ^{S}, in decreasing order.

ECF Sigma Factors

The ECF subgroup of alternative σ factors typically regulate functions related, in the broadest sense, to the cell envelope. Many bacterial species contain multiple ECF sigma factors, as shown in Table 6.4. Most ECF sigma factors are normally cotranscribed with one and rarely with more negative regulators. One includes a bitopic membrane protein where the C-terminal part is located outside of the cytoplasmic membrane, while the N-terminal part is exposed into the cytoplasm and is involved in binding its cognate sigma factor. Upon receiving a stimulus from the outside, the sigma factor is released and binds to the RNA polymerase core enzyme, stimulating transcription of those genes being under the positive control of the ECF sigma factor. The roles and mechanisms of regulation for these various ECF sigma factors are largely unknown. Two systems have been studied in detail, namely the σ^{E}-RseA and the σ^{W}-RsiW pairs, and these are described in detail under 9.2 and 9.8.2. While the former is involved in the heat-shock response of *E. coli*, the latter deals with alkaline stress in *B. subtilis*.

Table 6.4 Occurrence of ECF sigma factors in some bacterial species.

Species	*Number of ECF sigma factors*
E. coli	2 (involved in heat shock and iron uptake)
B. subtilis	7
C. diphtheriae	7
M. tuberculosis	10
C. crescentus	13
M. loti	16
P. aeruginosa	~19
S. coelicolor A3(2)	41

The NusA Protein

The NusA (for *N u*tilization *s*ubstance *A*) protein of *E. coli* binds to core RNA polymerase shortly after the initiation of transcription and stimulates pausing and termination at certain sites. The mechanism by which NusA influences pausing during transcription is not yet clear, but RNase protection experiments suggest that NusA may bind and stabilize the stem-loop RNA structures often associated with pause sites. By increasing the dwell-time of RNA polymerase at such a pause site, this stabilization may serve to couple transcription and translation. *In vitro* studies have shown that NusA also enhances termination of transcription at intrinsic terminators. These terminators also contain a GC-rich stem-loop in the nascent transcript upstream of the terminations site; and it is possible that NusA helps promote release of the transcript by stabilizing the RNA stem-loop and blocking its interaction with a single-stranded RNA-binding site on RNA polymerase. Consistent with this, the NusA in a transcription complex could be cross-linked to the nascent RNA located more than ten nucleotides from the 3′ end of the transcript. NusA binds directly to the α subunit of RNA polymerase. It can also be cross-linked to the large β and β′ subunits of RNA polymerase and may be capable of binding directly to these subunits as well. NusA influences not only pausing and termination by RNA polymerase, but also transcriptional anti-termination by the phage λ N protein (see below).

6.2.2 Regions of the Promoter

Promoters in bacteria contain recognition sequences for RNA polymerase in up to three distinct region: the –10 and –35 elements (–24 and –12 in the case of σ^{54}, see above) and the *UP element*. The first recognition element to be discovered and the most conserved is the –10 region, also called the Pribnow–Schaller box. On the nontemplate strand, this has the sequence *T*A*T*AA*T*, from –12 to –7, where the nucleotides in italics are ~80% conserved and the others ~60% conserved. The second recognition element is centered about 35 bp upstream of the first transcribed nucleotide and has the conserved sequence TTGACA. Besides these two recognition elements, the distance between the two elements has a characteristic length of 17±1 bp. This sequence is called the *spacer region* and, while sequence changes within that region hardly influence the promoter activity, altering the length has a dramatic effect on its activity by lowering it. Naturally occurring promoters diverge from the consensus sequence to varying degrees; and these variations are required to obtain physiological appropriate transcription rates. The role of the –10 region is complex, as it is recognized initially as duplex DNA and continues to function after the DNA is opened.

The UP Element

UP elements act as a third recognition element for the RNA polymerase, are located upstream of the –35 region and can enhance promoter strength significantly

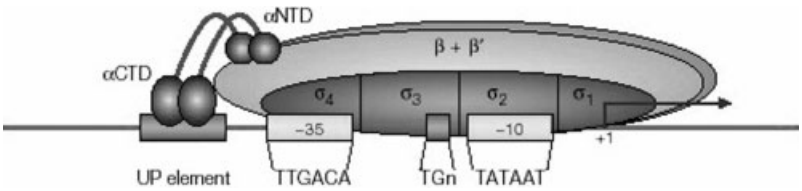


Fig. 6.7 The αCTDs of the α subunits interact with the UP element. The UP element is an AT-rich sequence located upstream of the –35 region of some promoters.It is recognized by the αCTDs, and their interaction with the UP element increases the promoter strength. D.F. Browning, S.J.W.Busby **2004**, *Nat. Rev. Microbiol.* 2, 1–9; Fig. 1b.

(Fig. 6.7). UP elements were demonstrated to be a docking site for the C-terminal domain of the α subunit of the RNA polymerase. They stimulate promoter activity primarily by increasing the initial equilibrium constant, although subsequent step(s) in the transcription initiation pathway may also be affected in at least a subset of UP element-containing promoters. UP elements are independent promoter modules because they can stimulate transcription when fused to other promoters. A full UP element, characterized by a high A+T content, actually consists of two subsites, each of which can bind an α subunit. The optimal (consensus) UP element sequence contains two parts, a proximal subsite, centered at about position –42 (consensus 5′-AAAAAARNR-3′), and a distal subsite, centered at about position –52 (consensus 5′-AWWWWWTTTTT-3′). The selected proximal subsites stimulate transcription up to 170-fold; and the selected distal subsites stimulate transcription up to 16-fold. Function of the consensus proximal subsite requires only one copy of αCTD, whereas function of the consensus distal subsite requires both copies of αCTD. It has been proposed that each subunit constitutes a binding site for a copy of αCTD and that binding of an αCTD to the proximal subsite region is a prerequisite for binding the other αCTD to the distal site. Analysis of the *E. coli* genome sequence suggests that UP elements, consisting of either one or two subsites with no more than two mismatches to consensus, occur in more than one-third of stable RNA promoters, but are also found in about 4% of mRNA promoters. Because sequences with much lesser degrees of match to consensus can still significantly affect promoter function, the number of promoters in which UP elements play some role in transcription is likely to be much larger. The first UP element was identified as a binding site for the α subunit of the RNA polymerase in the rRNA promoter of *rrnBP1* of *E. coli*, where it stimulates the promoter activity ~30-fold. UP elements have been identified in other promoters in *E. coli*, in other bacterial species and in promoters transcribed by holoenzymes containing alternative sigma factors.

The Extended –10 Motif

Activator-independent promoters recognized by *E. coli* E σ^{70} have been reported where the –10 region alone provides a sufficient signal for the initiation of tran-

scription and hence contacts with bases in the –35 hexamer are not essential. Transcription initiation at these promoters depends on two sequences: the –10 hexamer and the dinucleotide 5′-TG-3′ located at position –14/–15 relative to the transcription start site. These promoters were named ''extended –10'' promoters. This results in an extended –10 element, 5′-TGnTATAAT-3′, which appears to create alternative contact points for RNAP, most likely via the sigma subunit, such that transcription can be initiated in the absence of specific –35 region contacts. In support of this idea, it could be shown that RNA polymerase containing an altered form of sigma which lacks the C-terminal 84 amino acids (including region 4.2, which is responsible for contacting the –35 hexamer) could initiate transcription at a consensus extended –10 promoter (but which was unable to initiate transcription at a typical promoter with –10 and –35 regions resembling the consensus but without the 5′-TG-3′ extension). The extended –10 promoter element is present in ~20% of all *E. coli* promoters. In addition to the 5′-TG-3′ determinant, in Gram-positive bacteria there is extended sequence conservation to give the consensus 5′-TRTG-3′, known as the –16 region.

6.2.3 Initiation of Transcription

Transcription initiation is a multistep process and has been studied most extensively on *E. coli* σ^{70}-dependent promoters:

- $E\sigma^{70}$ binds to promoter DNA to yield an RNA polymerase-promoter closed complex (RP_{c1}) extending from approximately –55 to –5 relative to the transcription start site in which the DNA remains double-stranded.
- Extension of the RNA polymerase–DNA contacts downstream of the start site to about +20 characterizes the second closed complex (RP_{c2}).
- $E\sigma^{70}$ melts ~14 bp of DNA (transcription bubble; positions –11 to +3 relative to transcription start) to yield an RNA polymerase–promoter open complex (RP_o), a step called *isomerization*.
- $E\sigma^{70}$ enters into abortive cycles of synthesis and release of short RNA products as an RNA polymerase–promoter initial complex; and, ultimately:
- $E\sigma^{70}$ escapes the promoter (called *promoter clearance*) and enters into productive RNA synthesis as an RNA polymerase–DNA elongation complex.

Molecular Mechanisms Ensuring the Distribution of RNA Polymerase Between Competing Promoters

Most of the active RNA polymerase molecules are channelled into copying the genes encoding stable RNAs involved in translation, while other molecules are bound nonproductively to the cell's DNA. Therefore, the amount of free RNA polymerase molecules that is available to copy most of the about 4000 genes coding for proteins is short in supply. In addition, the amount of σ factors is limited, resulting in an intense competition between the different promoters of RNA poly-

merase. Five distinct molecular mechanisms ensure distribution of RNA polymerase between competing promoters. These involve: (a) promoter DNA sequences, (b) sigma factors, (c) small ligands, (d) transcription factors and (e) the folded bacterial chromosome.

Promoter Sequences

The RNA polymerase is faced with about 2000 different promoter sequences. Differences between these sequences largely influence the initiation frequency, with the promoters having near-consensus sequence elements functioning more efficiently. Many of the strongest bacterial promoters known have effective UP elements which function by binding the αCTD, as mentioned above. But these differences provide only static regulation that cannot be modulated according to environmental conditions. Therefore, most adaptive regulation is due to modulation by *trans*-acting factors.

Sigma Factors

As already mentioned, *E. coli* codes for one housekeeping sigma factor, σ^{70}, which allows RNA polymerase to recognize most promoters. In addition, the *E. coli* chromosome encodes six alternative sigma factors that accumulate in response to specific stresses. These alternative sigma factors compete with σ^{70} for the RNA polymerase core enzyme; and the holoenzyme is directed to a specific subset of promoters. The regulation of these alternative sigma factors can be very complicated, involving transcriptional, translational and post-translational control. This will be discussed elsewhere in this book. In many cases, the activity of an alternative sigma factor is controlled by an anti-sigma factor (see below).

Small Ligands

Small ligands allow RNA polymerase to respond quickly and efficiently to changes in the environment. One example is guanosine 3′,5′ biphosphate (ppGpp), which is synthesized when amino acid availability is restricted to an extent that results in limitation of translation. ppGpp, also called *alarmone* or *magic spot*, works by destabilizing open complexes at those promoters that control synthesis of the translation machinery. While the interaction of ppGpp with RNA polymerase is not promoter-specific, ppGpp-dependent inhibition only occurs at promoters that form unstable open complexes. Such promoters typically have short runs of GC-rich sequences near position +1; and they control many of the genes involved in translation. Such promoters are also unable to function well at low concentrations of the initiating nucleotide, usually ATP. It has been suggested that ppGpp controls expression in response to sudden starvation, whereas the internal ATP level controls expression in response to the growth rate. In other words, the open complexes of these promoters must be stabilized, requiring higher ATP and lower ppGpp concentrations.

Transcription Factors

The *E. coli* chromosome is predicted to code for more than 300 proteins binding to promoter regions and either up- or downregulating transcription. Most of these proteins are sequence-specific DNA-binding proteins targeted to specific promoters. It has been estimated that seven transcription factors (CRP, FNR, IHF, Fis, ArcA, NarL, Lrp) control 50% of all regulated genes, while ~60 transcription factors control only a single promoter. On the basis of sequence analysis, bacterial transcription factors can be grouped into different families; and a dozen families have been identified, among them the LacI, AraC, LysR, CRP and OmpR families.

Transcription factors couple the expression of genes to environmental signals, which means that they have to be regulated by controling either their activity or their expression. Four different mechanisms have been described so far to achieve this:

1. Modulation of the DNA-binding affinity of the transcription factor by small ligands, the concentration of which fluctuates in response to nutrient availability or stress. The most prominent example is the reduction of the DNA-binding activity of the Lac repressor by the small molecule allolactose, signaling the presence of lactose in the medium.
2. Modification of the activity of some transcription factors by covalent modification by phosphorylation as part of a two-component signal transduction system. In short, these systems consist of a sensor kinase, normally embedded in the cytoplasmic membrane, which, upon sensing the appropriate signal, first autophosphorylates itself and then transfers the phosphoryl group to the cognate response regulator, a transcription factor (see Section 6.2.9 for an extensive description of two-component systems).
3. The concentration of some transcription factors in the cell controls their activity. Here, the intracellular concentration is determined either by regulation of expression of the transcription factor or by proteolysis.
4. Sequestration of the transcription factor by a regulatory protein. Examples are anti-sigma factors (see Section 6.2.6) and anti-activators (see Section 6.2.8.2).

The Folded Bacterial Chromosome

Bacterial chromosomes are highly compacted by supercoiling and by interactions with proteins and RNA (see Section 2.2). In *E. coli*, a dozen so-called nucleoid proteins are involved in this compaction. The binding of these nucleoid proteins to DNA and the resulting folding in domains must affect the distribution of RNA polymerase between promoters, but our knowledge is still very limited. The best studied case is the H-NS protein, which completely silences gene expression at the promoters of the *proU* and *bgl* operons (see Sections 6.1.4 and 9.3.1, respectively).

A. Barnard, A. Wolfe, S. Busby **2004**, Regulation at complex bacterial promoters: how bacteria use different promoter organizations to produce different regulatory outcomes, *Curr. Opin. Microbiol.* 7, 102–108.

D.F. Browning, S.J.W. Busby **2004**, The regulation of bacterial transcription initiation, *Nat. Rev. Microbiol.* 2, 1–9.

6.2.4 Elongation of Transcription

Transcription elongation is initiated when the RNA polymerase leaves the promoter (promoter clearance). This leads to the release of the sigma factor from the complex and to a conformational transition of the enzyme, resulting in stable binding of the template DNA and a high processivity. But there are three situations where the elongating complex is modified, causing interruptions to transcription:

- Pausing, a temporary delay in chain elongation, synchronizes transcription and translation, slows RNA polymerase to allow timely interaction of regulatory factors and is a precursor to both arrest and termination.
- Arrest, a complete halting without dissociation, needs additional factors to resume transcription.
- Termination leads to dissociation of the transcription elongation complex at Rho-dependent and Rho-independent terminators.

Although *transcriptional pausing* was first described more than two decades ago, no consensus pause sequence exists. Rather, different types of signals appear to inhibit the alignment of the RNA 3′ OH with substrate NTP in different ways. Furthermore, numerous auxiliary proteins modulate pausing in organisms from bacteria to humans. Two of these, NusA and NusG, are universally conserved among bacteria and archaea, are typically essential to cell viability and inhibit or stimulate pausing by bacterial RNA polymerase.

In an *arrested state*, the RNA polymerase cannot continue RNA synthesis but also does not dissociate from the DNA template and RNA transcript. According to one model, arrest is associated with backward translocation of the RNA polymerase, which disengages the 3′ end of the transcript from the catalytic center of the enzyme. Factor-stimulated cleavage of an internal phosphodiester bond results in the generation of a new 3′ OH in register with the catalytic center, which allows for renewed RNA synthesis. Evidence for this model comes from experiments indicating that the RNA polymerase catalytic site also performs the cleavage reaction and that backtracking occurs for both eukaryotic and prokaryotic RNA polymerases. In eubacteria, the two transcript cleavage factors GreA and GreB have been identified, which both stimulate the intrinsic endonucleolytic activity of RNA polymerase. The factor-stimulated cleavage of RNA occurs 2–18 bases upstream from the 3′-terminus, followed by dissociation of the 3′-proximal fragment from the transcription elongation complex. The 5′-proximal fragment of the transcript remains in the transcription elongation complex and can be extended in the presence of rNTPs. GreA induces the hydrolysis of predominantly di- and trinucleotides (type A cleavage activity), whereas GreB induces cleavage of 2-nt to 18-nt RNA fragments (type B cleavage activity). In the absence of factors, the endonucleolytic activity of RNA polymerase can be induced by pyrophosphate or at alkaline pH, indicating that the same active center of RNA polymerase is responsible for the cleavage reaction, as well as for pyrophosphorolysis and RNA synthesis re-

actions. The precise molecular mechanism by which Gre factors activate the endonucleolytic activity of RNA polymerase is not known. The biological role for factor-induced endonucleolytic reactions may include:

- to enhance transcription fidelity by promoting excision of misincorporated nucleotides by RNA polymerase (transcription proofreading);
- to suppress transcriptional pausing and arrest by the 'cleavage and restart' mechanism during reversible and irreversible backtracking of transcription complexes;
- to stimulate RNA polymerase promoter escape and transition from initiation to elongation stage by suppressing early RNA release in favor of extension by the 'cleavage and restart' mechanism during abortive syntheses.

If the elongating RNA polymerase is arrested by a DNA damage (e.g., a pyrimdine dimer), another factor causes release of the stalled enzyme. This factor is the 130-kDa Mfd monomer (for *m*utation *f*requency *d*ecline), a DNA-dependent ATPase that requires ATP for release of RNA polymerase. Mfd includes a DNA-binding domain, an ATPase domain and an RNA polymerase–interaction domain. Mfd releases stalled RNA polymerase by binding immediately upstream of the TEC (*t*ranscription *e*longation *c*omplex) and then releasing RNA polymerase in an ATP-dependent reaction, perhaps through rewinding the DNA upstream of the transcription bubble. Mfd recruits the DNA excision repair machinery to the site of a DNA damage, initiating the repair of the DNA damage after RNA polymerase release. Mfd is a widely conserved bacterial protein that couples DNA repair with transcription.

Transcriptional Slippage

Transcriptional slippage involves the incorporation of nontemplated nucleotides into the mRNA by the elongating RNA polymerase. One documented case is the gene *dnaX* of *T. thermophilus*, encoding the τ and γ subunits of DNA polymerase III. While τ is the full-length product, γ results from a frameshifting event. Frameshifting is due to the RNA polymerase that introduces nontemplated nucleotides while transcribing a region of *dnaX* that contains nine A residues. During elongation, a hybrid of 8–9 nucleotides between the mRNA and the DNA template in the transcription elongation complex keeps the RNA polymerase in register with the transcribed sequence. The mechanism proposed for transcriptional slippage involves the dissociation of this hybrid and the repairing of the mRNA with the DNA in an offset position, which is possible over a series of identical nucleotides of sufficient length. When repairing occurs backwards, the RNA polymerase continues transcription by incorporating nucleotides as dictated by the template, resulting in mRNA that has additional nucleotides as compared to the template.

6.2.5 Termination of Transcription

In bacteria, two known transcription termination mechanisms allow dissociation of the RNA polymerase and its transcript from the DNA: intrinsic terminators and the Rho factor.

Intrinsic Terminators

Factor-independent terminators are also called *intrinsic terminators* and are characterized by two sequences: an inverted repeat as a G/C-rich (interrupted) palindromic region, followed by a trail of A residues on the template strand. The palindromic region is believed to be extruded as a hairpin in nascent RNA, causing the RNA polymerase to pause and weakening its interaction with the nascent RNA and template DNA. Final release is facilitated by the U-trail (7–8 U residues) probably due to the unusually weak nature of the rU-dA hybrid. This property, combined with the formation of a stem-loop structure in the RNA in the exit channel of RNA polymerase, is sufficient for the release of the transcript to dissociate from the complex before the next nucleotide is added. In addition, recent studies indicate that the primary role of the U-trail might be in stalling the polymerase and thereby providing time for the hairpin to form.

The Rho Factor

The only transcription termination factor known in bacteria is the Rho factor, which mediates release of the transcript at a DNA sequence at which the transcription complex is too stable for spontaneous release. It employs the energy from ATP hydrolysis to mediate the dissociation of a nascent transcript. In contrast to factor-independent terminators, Rho-dependent terminators are not definable by a consensus sequence. Factor Rho in *E. coli* consists of six identical, wedge-shaped protein subunits arranged in a gapped or closed ring structure (see below). A Rho monomer consists of 419 amino acid residues arranged in a N-terminal RNA-binding domain and an ATP-binding domain. Rho has a strong preference for single-stranded RNA rich in C residues which has a length of 40 or more nucleotides; and these binding sites are called *rut* (for *R*ho *ut*ilization sites). Besides *rut*, a Rho-dependent terminator consists of a second part, called the *tsp* (for *t*ermination *s*top *p*oints) region. This second part can extend for nearly 100 bp and determines the region where the transcripts are terminated at several positions, often in clusters of five to eight successive stop points separated by regions of 20–30 bp where few transcripts are terminated. The *tsp* regions contain naturally pause sites that are dictated by several features of the DNA sequence in and around the base pairs at which pausing occurs. The distance between *rut* where Rho loads on the RNA and where the transcript is terminated is relatively short, in the range 20–40 nucleotides.

Once bound, Rho acts as a hexameric, 5′→3′ ATP-dependent helicase to translocate to the site of transcription and disengage the RNA polymerase. Binding of Rho to a *rut* site is independent of ATP binding and/or hydrolysis and is mediated by a N-terminal domain in each monomer of the Rho hexamer. Rho can oscillate between closed- and open-ring states (Fig. 6.8, steps 1, 2), either of which may associate with ATP and the transcript (Fig. 6.8, steps 3, 4). Loading of the transcript onto the Rho hexamer occurs through the association between the *rut* site and the primary RNA binding sites in the N-terminal domains. If the ring is in the open conformation, the RNA threads through the gap, followed by closure of the ring. In contrast, if the ring is in the closed conformation, transcript entry must wait until the ring spontaneously opens. Next, the RNA molecule is sensed by the Q and R loops, leading to an allosteric activation of Rho and thereby converting it into a stable, closed configuration (Fig. 6.8, step 5). The last step involves translocation of the transcript powered by ATP hydrolysis. The translocation involves either dissociation of Rho from the *rut* site or, alternatively, it could remain bound to the target motif and track along the mRNA (Fig. 6.8, steps 6, 7). Rho can be regulated during this process by a number of factors, such as NusG, which serve as anti-terminators to ameliorate the termination properties of the enzyme.

The Rho factor is not present in all eubacteria as, e.g., *Mycoplasma* and *Synechocystis*, or it is present but not essential as, e.g., in *B. subtilis* and *S. aureus*. What could be the selective advantage for a cell to carry a *rho* gene? One likely advantage is that, under conditions of stress, the Rho protein prevents the continued synthesis of an RNA unable to become translated. One example is the *lac* operon which contains latent Rho-dependent intragenic terminators. In the absence of an amino acid, movement of the ribosomes is slowed or blocked and a *rut* segment becomes

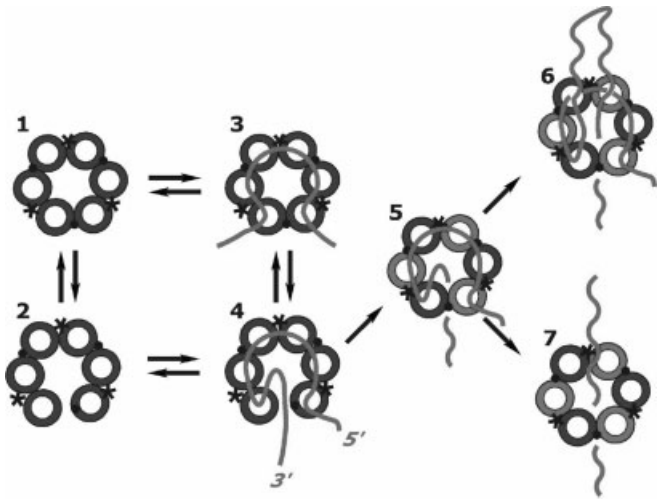


Fig. 6.8 Model of factor Rho function. See text for detailed explanation.
E. Skordalakes **2003**, *Cell* 114, 135; Fig. 5.

exposed, allowing Rho to bind to the mRNA and to terminate the partial transcript.

T. Henkin **1996**, Control of transcription termination in prokaryotes, *Annu. Rev. Genet.* 30, 35–57.
J.P. Richardson **2002**, Rho-dependent termination and ATPases in transcript termination, *Biochem. Biophys. Acta* 1577, 251–260.

6.2.6 Anti-sigma Factors

Anti-sigma factors are proteins which can interact with sigma factors to prevent their binding to the RNA polymerase core enzyme. This interaction can occur transiently or permanently. Several anti-sigma factors have been characterized in the past few years.

Anti-sigma Factors Interacting with the *E. coli* σ^{70} Housekeeping Factor

The first demonstration of an inhibitor protein binding to sigma emerged from studies of RNA polymerase modification during virulent phage T4 infection. This anti-σ factor, AsiA (for *a*nti-*si*gma factor *A*), associates tightly with *E. coli* σ^{70} and inhibits transcription from both host promoters and early, σ^{70}-dependent phage promoters. The small 10.6-kDa AsiA protein binds in a 1:1 complex with σ^{70} and thereby blocks recognition of the –35 promoter region. Amino acid residues 567–600 of σ^{70} are sufficient to bind AsiA, suggesting that this region contains the AsiA-binding site. This stretch of amino acid residues also defines σ^{70} domain 4.2 and includes residues that make direct contact with the –35 region present in host and T4 early promoters. The binding of AsiA specifically inhibits transcription from promoters that require a contact between domain 4.2 and the –35 region of a promoter, but does not affect transcription from extended –10 promoters, which do not require this contact. These observations led to a simple model in which the interaction of domain 4.2 in σ^{70} with the –35 promoter element or with AsiA are mutually exclusive. The three-dimensional structure of an AsiA/region 4 complex revealed that the C-terminal α helix of region 4 is unstructured, while four other helices adopt a completely different conformation relative to the canonical structure of unbound region 4. Therefore, AsiA induces a rearrangement in region 4, preventing its interaction with the –35 region of σ^{70}-dependent promoters.

Biochemical fractionation of σ^{70}-associated proteins from stationary phase cells identified a single, specifically associated polypeptide termed Rsd (for *r*egulator of *s*igma *D*). This protein binds to region 4 of σ^{70} and appears to block association with RNA polymerase core enzyme. Since the levels of Rsd *in vivo* are only sufficient to complex 20% of σ^{70}, the role of this protein is not immediately obvious. Rsd may facilitate the replacement of σ^{70} by the stationary phase-specific sigma factor σ^{S} in functional RNA polymerase holoenzyme complexes as cells go from the exponential phase to the stationary phase of growth.

Anti-sigma Factor Regulating Flagellar Gene Expression

Bacteria use flagella to move away from stressful areas into microenvironments favorable for growth. The synthesis of late flagellar genes encoding flagellin and chemotaxis functions depends on the alternate sigma factor σ^F. During the early stages of flagellar biogenesis, σ^F is held inactive by tight association with FlgM. Once the hook and basal body are assembled, FlgM is secreted by the flagellar export system, thereby freeing active σ^F. FlgM can be exported from the cell by passing through the hollow inner core of the basal body, hook and flagellar filament structures. The mechanism by which FlgM inhibits the activity of the σ^F RNA polymerase involves both the sequestration of σ^F and the destabilization of existing σ^F holoenzyme.

Anti-sigma Factors Interacting with *E. coli* Heat-shock Sigma Factor

Subunit σ^E (σ^{24}) is a member of the ECF family of sigma subunits for transcription of the genes for proteins involved in extracytoplasmic functions, as well as those required for survival at high temperatures. The σ^E activity is regulated by the *rseA* (for *r*egulator of *s*igma *E A*) gene product, the anti-sigma factor which is anchored in the inner membrane and inhibits the activity of σ^E by direct interaction (see Section 9.2 for a detailed description).

Regulation of Anti-sigma Factors by Partner-switching Modules

Modulation of sigma factor activity can also occur by an *anti-anti-sigma factor*, where the anti-sigma can bind either to the cognate sigma factor or to the anti-anti-sigma factor. This control circuit where the anti-sigma factor can interact with one of two other proteins has been designated as the partner-switching module; and two such modules have been described for *B. subtilis* so far. One involves the sporulation-specific sigma factor σ^F and the other involves the general stress factor σ^B. Under physiological conditions, both sigma factors form a complex with their cognate anti-sigma factor (Fig. 6.9). These anti-sigma factors also act as kinases which phosphorylate their cognate anti-anti-sigma factors, thereby keeping them inactive. Upon an appropriate signal arising in the cytoplasm, the anti-anti-sigma factors are dephosphorylated by a phosphatase, attack the complex consisting of the anti-sigma factor and sigma factor, bind to the anti-sigma factor and cause release of the sigma factor.

Regulation of an Anti-sigma Factor by Redox Change

Another recently discovered principle makes use of the redox potential within the cytoplasm. The extracytoplasmic function sigma factor, σ^R, is required for induction of the thioredoxin reductase/thioredoxin operon (*trxBA*) of *S. coelicolor* in response to various oxidants, including the thiol-specific oxidant diamide. σ^R also directs diamide-inducible transcription from one of the promoters of its own synthesis. The transient induction *in vivo* of *trxBA* transcription upon diamide treat-

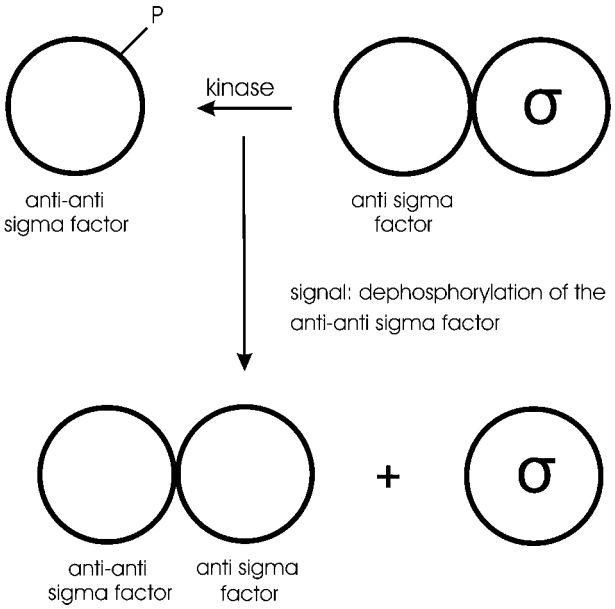


Fig. 6.9 The partner switching model of anti and anti-anti sigma factors. Under physiological conditions, the sigma factor phosphorylates its cognate anti-anti sigma factor and forms a stable complex with its sigma factor. If the appropriate signal appears in the cytoplasm, dephosphorylation of the anti-anti sigma factor occurs which binds to the preexisting complex and causes release of the sigma factor.

ment suggests a model in which σ^R induces expression of the thioredoxin system in response to cytoplasmic disulfide bond formation. Upon reestablishment of normal thiol levels, σ^R activity is switched off, resulting in downregulation of *trxBA* and *sigR*. However, since σ^R itself contains no cysteines, there must be another component in the system to explain how σ^R activity responds to thiol oxidation and reduction. The activity of σ^R is modulated by a specific anti-sigma factor, RsrA, which binds σ^R and inhibits σ^R-directed transcription *in vitro* only under reducing conditions. Exposure to H_2O_2 or to the thiol-specific oxidant diamide causes the dissociation of the σ^R-RsrA complex, thereby allowing σ^R-dependent transcription. This correlates with intra-molecular disulfide bond formation in RsrA. Thioredoxin is able to to reduce oxidized RsrA, suggesting that σ^R, RsrA and the thioredoxin system comprise a novel feedback homeostasis loop that senses and responds to changes in the intracellular thiol-disulfide redox balance.

6.2.7
Anti-termination/Attenuation of Transcription

Bacteria have evolved sophisticated mechanisms to control both transcription and translation of genes in response to environmental changes. While in many cases, transcription is controlled at the level of initiation by DNA-binding proteins (see

below), the transcription elongation complex is also a potential target for regulation where the nascent transcript may fold into secondary structures causing the RNA polymerase to pause or even to terminate transcription. Two different mechanisms have been described, designated *transcription attenuation* and *transcription anti-termination*. In transcription attenuation, the default pathway is readthrough and a regulatory molecule induces transcription termination. In transcription anti-termination, the default pathway results in premature termination and a regulatory molecule promotes transcription readthrough. It modifies the transcription apparatus in such a way that terminators can be bypassed. Both mechanisms involve a wide variety of *trans*-acting factors that interact with the mRNA to change its secondary structure to either favor or prevent formation of a transcription termination structure. In the following sections, six examples will be presented to illustrate attenuation and anti-termination: transcription attenuation of the *trp* operons of *E. coli* and *B. subtilis* and transcription anti-termination control by the phage λ proteins N and Q, by the BglG/SacY family of anti-terminator proteins and by tRNAs in Gram-positives.

Initiation of Transcription of the *E. coli trp* Operon is Negatively Regulated by the TrpR Repressor and Elongation is Fine-tuned by a Transcription Attenuation Mechanism

The first transcription attenuation mechanism was described for the *E. coli* tryptophan biosynthetic operon (*trpEDCBA*) and it was later shown that many amino acid biosynthetic operons in enteric bacteria are subject to transcription attenuation. Transcription initiation of the *trp* operon is regulated by TrpR, a DNA-binding repressor protein. In the presence of a high tryptophan concentration, this amino acid acts as a corepressor by binding to TrpR, thereby allowing this complex to interact with the operator sequence present within the promoter region. Binding of the TrpR-Trp complex to the operator prevents initiation of transcription. Once transcription has been initiated, the elongation complex is subject to transcription attenuation. The combined actions of repression (60-fold) and attenuation (8-fold) allow an about 500-fold regulation in response to the intracellular tryptophan concentrations. The *trp* mRNA starts with a 141-nucleotide leader transcript able to form three overlapping RNA secondary structures designated as pause structure, anti-terminator and intrinsic transcription terminator (Fig. 6.10). The leader transcript contains a small open reading frame coding for a 14-amino-acid leader peptide. Shortly after transcription has been initiated, the pause structure forms which signals the RNA polymerase to stop for a moment. During that pause time, a ribosome initiates translation of the leader peptide and disrupts the paused RNA polymerase complex, allowing the enzyme to resume transcription. The further fate of the transcribing complex is dictated by the amount of tryptophan present within the cell. If the level of this amino acid is low, the amount of aminoacylated $tRNA^{Trp}$ is also low. This in turn leads to the stalling of the translating ribosome at one of two tandem Trp codons present within the leader sequence. Ribosome stalling uncouples transcription and translation, leading to the

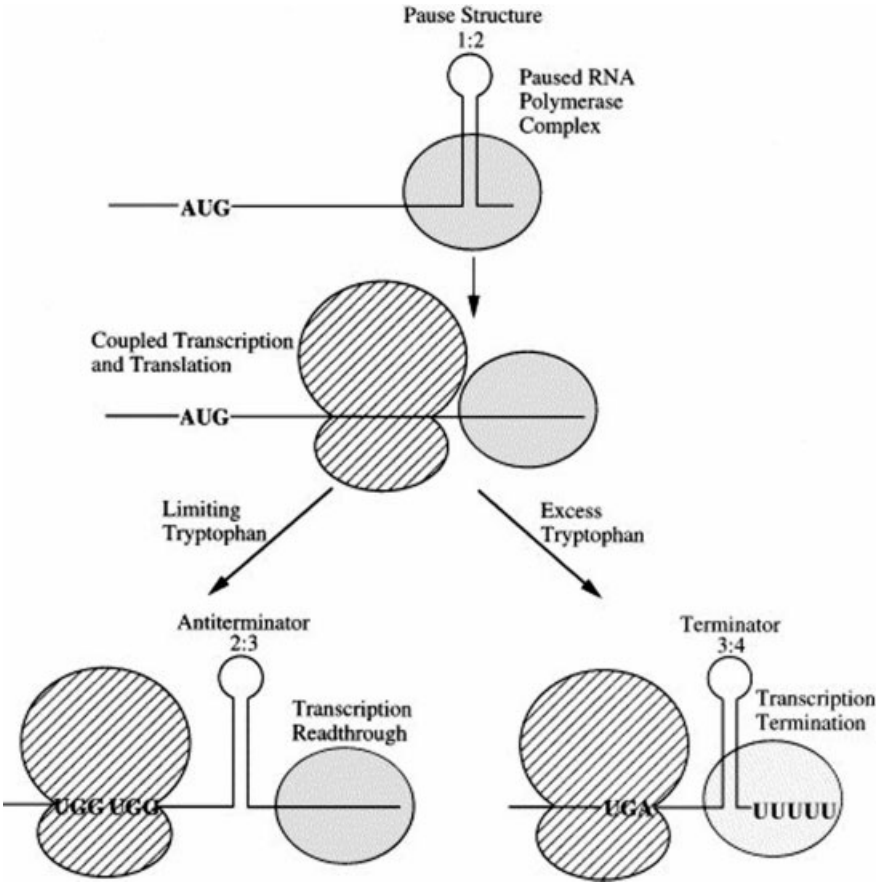


Fig. 6.10 Transcription attenuation of the *E. coli trp* operon is mediated by alternative secondary structures. Transcribing RNA polymerase stops at the pause structure 1:2. If the tryptophan concentration within the cells is limiting, ribosomes translating the leader sequence are stalled at two UGG codons due to the absence of aminoacylated $tRNA^{Trp}$ molecules leading to the formation of the anti-terminator structure 2:3. If the tryptophan concentration is high, the translating ribosomes reach the UGA stop codon resulting in the formation of the terminator structure 3:4. P. Gollnick **2002**, *Biochim. Biophys. Acta* 1577, 240; Fig. 1.

formation of the anti-terminator structure. This secondary structure prevents the formation of the terminator structure and transcription proceeds towards the end of the operon. If the tryptophan concentration is high, the ribosome is able to translate the tandem Trp codons and to reach the end of the reading frame, thereby blocking formation of the anti-terminator structure. This in turn promotes formation of the terminator hairpin and hence termination of transcription of the *trp* operon. In this attenuation mechanism, the charged $tRNA^{Trp}$ acts as the regulatory signal. Besides the *trp* operon, other amino acid biosynthetic operons such as

the *leu*, *his* and *phe* operons in enteric bacteria are regulated by a similar attenuation mechanism.

Expression of the *trp* Operon of *B. subtilis* is Regulated in Response to Tryptophan by the *trp* RNA-binding Attenuation Protein TRAP

The transcription attenuation mechanism regulating the *trp* operon in *B. subtilis* and other Gram-positive bacteria is completely different from that described for *E. coli*. Instead of translating ribosomes, a sequence-specific RNA-binding protein, TRAP (*trp R*NA-binding *a*ttenuation *p*rotein), senses the level of tryptophan within the cell (Fig. 6.11). Transcription of the *trpEDCFBA* operon of *B. subtilis* starts 203 nucleotides upstream from the first structural gene constitutively and is then regulated by attenuation. In agreement with the *E. coli trp* leader region, this leader region contains several inverted repeats, too, capable of forming mutually exclusive anti-terminator and intrinsic terminator RNA secondary structures. In addition, there is a stem-loop structure at the 5′ end of the transcript. When a high concentration of tryptophan is present within the cell, the attenuation protein TRAP binds to a series of 11 trinucleotide NAG repeats (GAG > UAG > AAG > CAG), which are separated from each other by several nonconserved spacer nucleotides. These trinucleotide repeats are present between +36 and +91 in the leader transcript and overlap the 5′ portion of the anti-terminator structure (see Fig. 6.11). Upon the binding of TRAP, it prevents formation of the anti-terminator and thereby favors formation of the downstream terminator loop, which in turn results in transcription termination before the RNA polymerase has reached the first structural gene. TRAP binding can downregulate transcription of the strucural gene region of the *trp* operon, reducing it by ~200 times. TRAP binding also regulates tryptophan synthesis in a second way. It promotes formation of an additional RNA hairpin structure in the small fraction of leader transcripts that are not terminated at the attenuator. This hairpin structure inhibits *trpE* translation, further reducing the ability of the cell to synthesize tryptophan. Under conditions of low tryptophan concentration, TRAP does not bind to the leader RNA, the anti-terminator forms allowing transcription readthrough into the *trp* operon. TRAP consists of 11 identical subunits, composed of 75 amino acids arranged in a ring-like structure. Activation of TRAP occurs by the binding of up to 11 molecules of L-tryptophan into the hydrophobic clefts between adjacent subunits. A total of 200–400 TRAP 11-mer molecules per cells have been determined.

Activation of transcription occurs by two mechanisms: (a) uncharged $tRNA^{Trp}$ accumulates leading to the activation of the *at* operon coding for an Anti-TRAP protein designated AT and encoded by the *rtpA* gene and (b) TRAP protein not loaded with tryptophan is unable to bind to the transcript. Activation of the *rtpA* gene occurs at the level of transcription and translation. The leader transcript of the *at* operon can fold to form alternative anti-termination/termination structures that are regulated by the T box transcription/anti-termination mechanism (see below). This mechanism involves the binding of uncharged $tRNA^{Trp}$ to the leader RNA, stabilizing an anti-terminator structure, which prevents formation of the

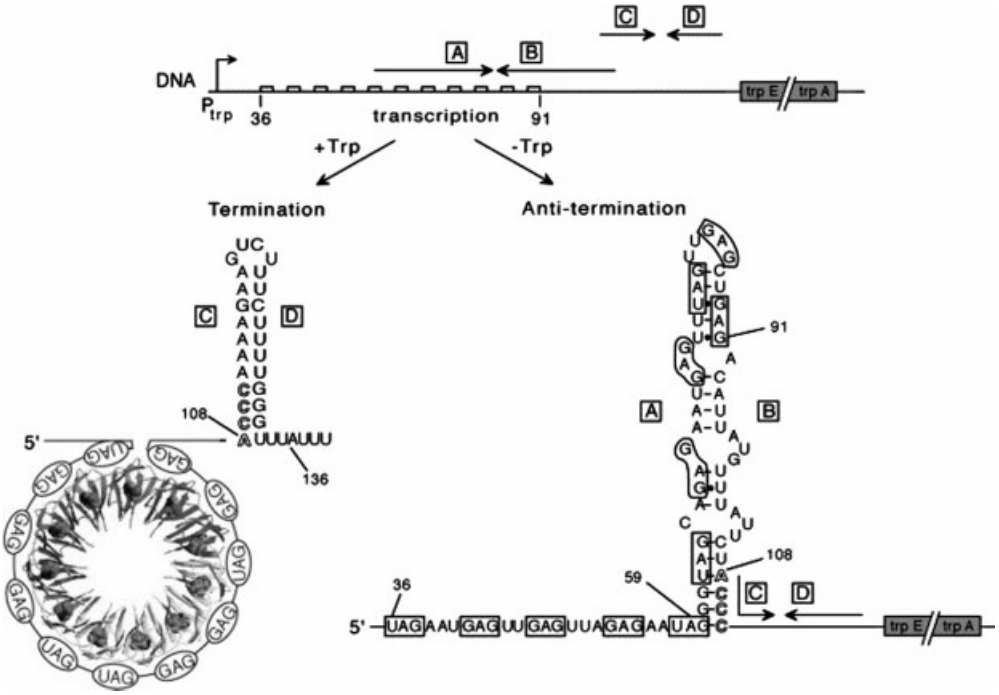


Fig. 6.11 Transcription attenuation of the *B. subtilis trp* operon is mediated by the RNA-binding protein TRAP. In the presence of a high tryptophan concentration within the cell, the TRAP protein is loaded with tryptophan and then able to bind to the leader region of the constitutively initiated *trp* transcript. This causes formation of the terminator structure C:D and attenuation of transcription. If the tryptophan concentration is low, TRAP will be unable to bind this amino acid resulting in no binding to the nascent transcript favouring formation of the anti-terminator structure A:B. P. Gollnick **2002**, *Biochim. Biophys. Acta* 1577, 240; Fig. 2.

terminator, allowing transcription to continue into the structural genes of the operon. Regulation at the level of translation occurs at the *rtpLP* gene coding for a ten-residue leader peptide. This coding region contains three consecutive Trp codons and its stop codon is located only six nucleotides upstream from the Shine–Dalgarno (SD) sequence of *rtpA*. This proximity suggests that the ribosome completing translation of *rtpLP* could inhibit initiation of *rtpA* translation, limiting AT synthesis. If the presence of charged $tRNA^{Trp}$ is low, the ribosome translating *rtpLP* stalls at one of the three Trp codons, exposing the *rtpA* SD sequence and permitting efficient AT synthesis. The AT produced binds to tryptophan-activated TRAP only, inhibiting TRAP function and allowing transcription of the *trp* operon. In the second mechanism, limiting amounts of tryptophan lead to empty TRAP proteins unable to interact with the *trp* leader sequence, preventing formation of the terminator structure. AT exists in a reversible equilibrium between trimer and dodecamer.

The Anti-terminator Proteins N and Q of Bacteriophage λ

Anti-termination in λ is induced by two quite distinct mechanisms. The first is the result of interaction between λ N protein and its targets in the early phage transcripts; and the second is the result of an interaction between λ Q protein and its target in the late phage promoter. When λ DNA enters the *E. coli* cell, it becomes circularized spontaneously, the four nicks are closed by the DNA ligase and negative supercoils are introduced by the DNA gyrase. Next, transcription of the so-called immediate early regions is initiated by the *E. coli* RNA polymerase at the two promoters P_L and P_R (Fig. 6.12). Most of the transcripts are terminated at the two transcriptional terminators t_L^1 and t_R^1, resulting in monocistronic transcripts in both cases. These two transcripts are translated into the proteins N and Cro (inhibitor of repressor synthesis), where N acts as an anti-terminator. N is a small basic protein of the arginine-rich motif family of RNA binding proteins which bind to the RNA sequence termed *nutL* (for *N ut*ilization *l*eftward) and *nutR* (*N ut*ilization *r*ightward). Both *nut* sites consist of two sequences called *boxB* and *boxA*. *boxB* forms a 15-nucleotide stem-loop structure in the mRNA which acts as a binding site for the N protein. *boxA* is the loading site for the *E. coli* NusB (*N u*tilization *s*ubstance *B*) and NusE (S10) proteins. In association with two additional factors, NusA and NusG, an anti-termination complex is formed with RNA polymerase. While N is the active factor in the complex, the additional components increase the stability of the complex. Binding of the different proteins occurs in a sequential manner. After the binding of N to *boxB*, NusA binds which requires extrusion of the loop residues of *boxB* forming the core complex. The assembly of NusB, NusE and NusG onto the core complex involves *boxA* and the C-terminal region of N. The N protein in the anti-termination complex converts the RNA

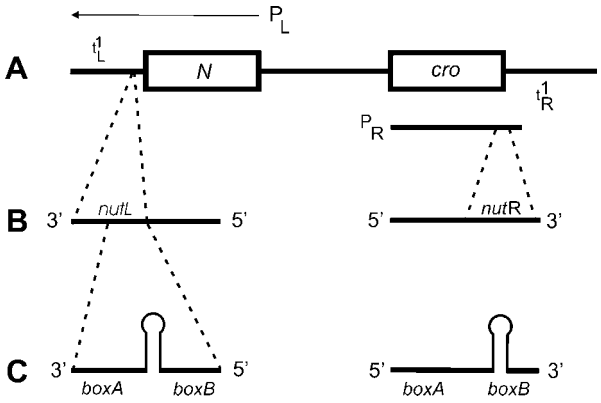


Fig. 6.12 Anti-termination mediated by bacteriophage λ protein N. (A) *E. coli* RNA polymerase will bind to the two promoters P_L and P_R resulting in the expression of the two genes N and *cro*. Most transcripts are terminated at t_L^1 and t_R^1. (B) The N anti-terminator protein will interact with *nutL* and *nutR*, recruit some *E. coli* Nus proteins and alter the conformation of the RNA polymerase is such a way that it will be able to override both intrinsic and Rho-dependent promoters. (C) The *nut* sites consist of two subsites called *boxA* and *boxB*.

polymerase in a termination-resistant enzyme; and anti-termination can still be detected after the polymerase has read through thousands of base pairs and many sequential terminators. This finding implies that the N-dependent modification to the polymerase that occurs after transcription of *nutL* and *nutR* is stable.

The protein Q induces anti-termination in the λ late operon, including the head and tail genes. It interacts with DNA, with the σ^{70} initiation subunit and with RNA polymerase core subunits. Its role in phage growth is to modify transcription initiated at the P_R' promoter of a single long transcript for phage late genes and, particularly, to extend transcription beyond a terminator immediately downstream of this promoter. Initially, Q binds to a DNA sequence called *qut* (*Q ut*ilization) consisting of a DNA binding element, located between the –35 and –10 promoter elements, and a segment downstream of +1 containing a promoter –10 element-like sequence. This sequence captures elongating RNA polymerase 16–25 nucleotides from the RNA start site in a σ^{70}-induced paused state that is primed for Q recognition. During this pause, Q binds DNA just upstream of the elongating enzyme, contacts region 4 of σ^{70} and displaces it forward to a promoter –35 element located between the –2 and –7 positions of the promoter. The pausing of the RNA polymerase at +16 is induced by interactions between σ^{70} and a DNA sequence with similarity to the –10 extended region (see above). This interaction occurs between Q and σ^{70}. How Q modifies the RNA polymerase is unknown. Sometimes, the RNA polymerase overrides the pause site and makes transcripts of ≥17 nucleotides in length. If this occurs, the RNA polymerase must backtrack by 16 nucleotides, while the GreA and GreB proteins stimulate an endogenous RNA polymerase ribonuclease activity that cleaves the extra RNA that extrudes from the RNA polymerase as a result of backtracking. In the absence of Q, the transcript is terminated at t_R'.

The *E. coli rrn* operons are regulated by an anti-termination mechanism that is dependent on sites that are closely related to λ *boxA*, has a promoter located proximal to the 16S and 23S structural genes in each operon and needs at least NusB, while the role of the other Nus proteins remains elusive.

Anti-termination of the *E. coli bgl* Operon Through the Protein BglG

Transcription of several catabolic operons in bacteria is regulated by anti-termination through an RNA-binding protein. This protein interacts with the transcript and prevents formation of the anti-terminator in nascent mRNA upstream of the structural genes. One of these systems in *E. coli* is the *bglGFB* operon, which codes for the uptake and utilization of aromatic β-glucosides. As already mentioned, this operon is cryptic in wild-type cells and can become activated through a mutation (see Section 6.1.5). The anti-terminator protein is encoded by the *bglG* gene and, in the absence of inducing sugar, BglG is inactive and does not bind to mRNA. Under these conditions, most transcripts are terminated at the Rho-independent terminators flanking *bglG* (Fig. 6.13A, B). In the presence of a β-glucoside, BglG homodimers bind to an RNA sequence called RAT (for *r*ibonucleic *a*nti-*t*erminator) partially overlapping with the terminator structure enabling the

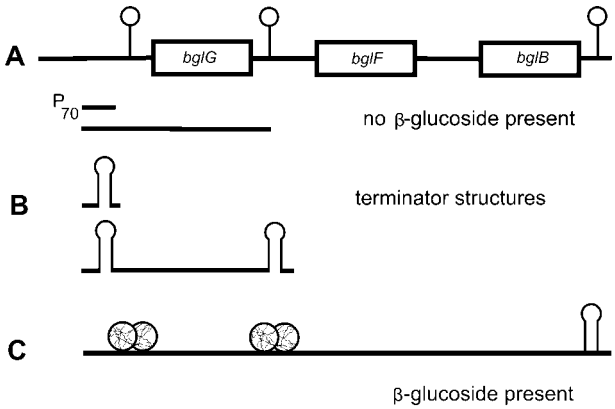


Fig. 6.13 Anti-termination of the *bgl* operon. (A) The *bgl* operon codes for the anti-terminator protein BglG, the β-glucoside permease BglF and the phospho-β-D-glucosidase BglB. Transcription starts at a σ^{70}-dependent promoter and is terminated at either of two intrinsic terminators flanking *bglG* in the absence of β-glucosides. (B) The two transcripts form one or two stem-loop structures, respectively. (C) If β-glucosides are present in the medium, the homodimeric BglG anti-terminator binds to the mRNA to prevent further formation of the transcriptional terminators.

RNA polymerase to proceed (Fig. 6.13C). How is the activity of BglG modulated in response to β-glucosides? This occurs through phosphorylation of BglG and involves the BglF protein, an enzyme II of the PTS pathway. This protein is inserted into the inner membrane and serves as a permease for β-glucosides. In the absence of these sugars, it recruits BglG to the membrane and phosphorylates it, preventing it from dimerization, a prerequisite for binding to RAT. In the presence of β-glucosides, the sugar molecules are phosphorylated and BglG is dephosphorylated by BglF. BglG (and the other members of this family of anti-terminator proteins) is composed of three domains: an RNA-binding domain, followed by two reiterated domains designated PRD1 and PRD2 (for *P*TS *r*egulation *d*omains). His-208 in the PRD2 domain is phosphorylated by BglF. Formation of BglG dimers initiates with PRD2 dimerization, followed by zipping up of two BglG monomers to create the active RNA binding domain.

Regulation of Aminoacyl-tRNA Synthetase Genes in Gram-positive Bacteria by Anti-termination

A completely different mechanism of anti-termination has been described for a large number of aminoacyl-tRNA synthetase genes in Gram-positive bacteria involving tRNA as the regulatory molecule. Expression of these genes is induced when the level of the cognate amino acid is low. The transcript of these aminoacyl-tRNA synthetase genes contain a long (about 300 nucleotides), untranslated leader region upstream of the structural gene, which can adopt a rather complicated secondary structure with three stem loop structures preceding the termina-

tor (Fig. 6.14). The regulatory tRNA molecule is able to interact with two sequence motifs within the secondary structure of the leader region, involving its anticodon and its CCA end. While the anticodon interacts with a triplet corresponding to the codon for the appropriate amino acid exposed on a bulge present within the 5′ proximal stem-loop (termed the *specifier sequence*) , the uncharged CCA ends with the complementary UGG sequence present in a bulge in the anti-terminator structure. This UGG sequence is part of a conserved 14-nucleotide sequence present in all leader sequences of aminoacyl-tRNA synthetase genes regulated by this mechanism, has been termed T-box and all these genes form the T-box family. Uncharged tRNA molecules signal starvation for the corresponding amino acid and, by binding to the leader sequence, they favor the formation of the anti-terminator and readthrough of the RNA polymerase. In the presence of high levels of the corresponding amino acid, all tRNA molecules are charged and their CCA ends are unable to interact with the anti-terminator, favoring formation of the transcription terminator structure.

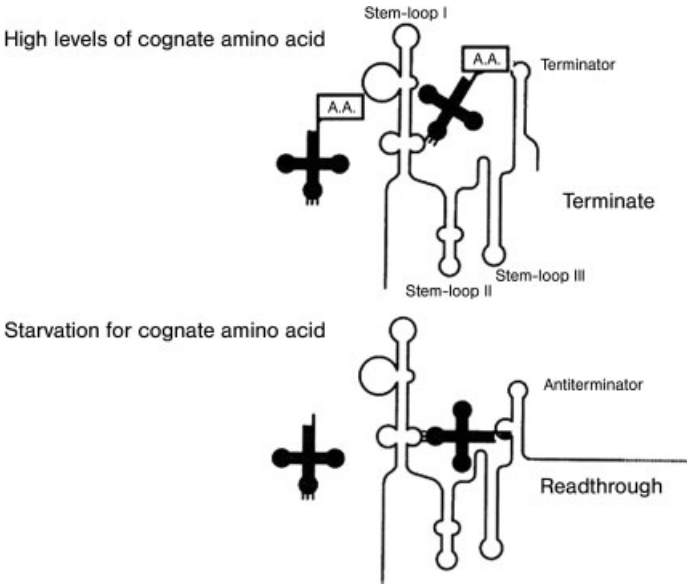


Fig. 6.14 Regulation of aminoacyl-tRNA synthetase genes in Gram-positive bacteria by anti-termination. The principle of termination and anti-termination is based on whether the cognate tRNA molecule is aminoacylated or not which in turn is dictated by the level of the amino acid present within the cell. Using its anticodon, the tRNA binds to a nucleotide sequence present within a bulge of stem-loop I. If the level of the cognate amino acid is high, the tRNA is aminoacylated and unable to interact with a bulge exposed on stem-loop III; this leads to the formation of the terminator structure. If the level is low, the uncharged tRNA will interact with this bulge favoring the formation of the anti-terminator and thereby readthrough into the structural genes.
P. Gollnick **2002**, *Biochim. Biophys. Acta* 1577, 240; Fig. 4.

6.2.8
Regulators of the DNA-dependent RNA Polymerase

Of the 71 *E. coli* transcription factors with experimentally verified regulatory information, 18 are activators, 20 are repressors and 33 are *dual regulators*. Dual regulators act both as repressors and as activators, depending on the location of the DNA binding site. Most of these factors are two-domain proteins, the remaining ones three- and four-domain proteins. Each transcription factor consists of one or two DNA binding domains and most have one or more dimerization domains. Specific binding to the DNA is accomplished in most cases by a helix-turn-helix motif, where one helix specifically recognizes the appropriate DNA sequence (the *recognition helix*), followed by the turn consisting of three amino acids where the first is normally a glycine. The second helix stabilizes the recognition helix through hydrophobic interactions. Most DNA-binding proteins are active as dimers, a few as monomers or tetramers. What determines whether a transcription factor acts as an activator or a repressor? A careful analysis of the binding sites for these 71 transcription factors has shown that the position of the binding site on the DNA relative to the transcription start site is indicative of its regulatory function. While the large majority of activators have only binding sites located upstream of the promoter, most of the repressor binding sites are downstream, or upstream in conjunction with a downstream site. Activators function by stabilizing the RNA polymerase from upstream sites and repressors act in most cases by steric hindrance, blocking polymerase binding or processing. Alternatively, transcriptional repressors block transition from a closed to an open complex or by inhibiting promoter clearance (see below for details).

Often, two adjacent transcriptional units are transcribed in opposite directions and separated by a short regulatory region containing at least the two promoters and sometimes binding sites of transcriptional regulators. These arrangements are designated *divergons* (Fig. 6.15). As to the location of the two promoters, there are three different possibilities: back to back, overlapping or face to face. In the first case, RNA polymerase can bind to the two promoters without interfering which each other. In the second case, RNA polymerase can interact with only one of the two promoters at a time. And in the third case, the two promoters collide if

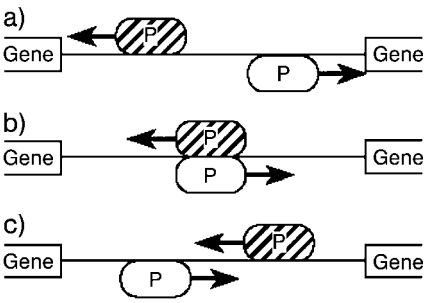


Fig. 6.15 Arrangements of promoters in divergons: (a) back-to-back, (b) overlapping and (c) face-to-face. M Yamada, et al. **2003**, *J. Mol. Microbiol. Biotechnol.* 6, 206–210; Fig. 1.

transcription starts at both locations. In most cases, the regulatory region between the two divergent operons contains binding sites for transcriptional regulators.

6.2.8.1 Transcriptional Repressors

Although repressors are generally believed to work by binding to the promoter in a way that impedes subsequent binding of RNA polymerase, the detailed analysis of several promoters has shown that *steric hindrance* is but one of the several mechanisms used by repressors to achieve their function.

6.2.8.1.1 Mechanisms of Repressor Action

Repressors Inhibiting RNA Polymerase Binding to the Promoter: Repression by Steric Hindrance Inhibition of RNA polymerase binding to a promoter can be achieved by binding a repressor protein to the promoter in a way that impedes RNA polymerase binding. Several repressors have been shown to work in this way, such as the phage λ cI repressor when binding to the O_{R1} and O_{L1} operator of the viral P_R and P_L promoter, the LexA and LacI repressors. To exclude RNA polymerase from the promoter, a repressor does not necessarily need to bind to a site overlapping the RNA polymerase binding site. The CytR repressor binds to position –70 at the *E. coli deo* promoter, helped by two flanking cAMP receptor protein (CRP) dimers bound to positions –40 and –93. The repressor is stabilized at the DNA by direct contacts with the two CRP dimers. The complex formed does not overlap the RNA polymerase binding site. The nucleoprotein complex formed by CytR and the two CRP dimers seems to wrap DNA in a way that impedes RNA polymerase binding to the promoter. A final interesting case is that of the abundant nucleoid-associated protein H-NS from *E. coli*, which is known to downregulate the expression of several genes. Despite having low sequence specificity, it binds preferentially to DNA regions showing intrinsic curvature, where it can form multimeric nucleoprotein complexes in which the DNA wraps around H-NS. These complexes can repress transcription in two ways: either physically blocking access of RNA polymerase to the promoter or altering the topology of DNA in the vicinity of the promoter, thus impairing transcription initiation. It seems that silencing of the *E. coli bgl* promoter is an example of the first possibility (the H-NS binding site covers the promoter), while repression of the *E. coli proU* promoter could be an example of the second (H-NS binds downstream from the promoter).

Repressors Blocking Transition From Closed to Open Complexes Several promoters have been shown to bind DNA in a way that allows simultaneous binding of RNA polymerase to the promoter, at least *in vitro*. In some cases, it has been published that, in such a ternary complex, RNA polymerase is unable to open the DNA strands at the –10 region and cannot proceed towards an open complex. We can distinguish two groups of repressors:

- those binding to sites overlapping the RNA polymerase binding region (about –40 to +10 relative to the start site), which may or may not form stable ternary complexes *in vivo* despite being able to form them in vitro;
- those binding to sites which do not overlap with RNA polymerase and which should have no problems binding close or adjacent to the RNA polymerase.

Among the first group, *B. subtilis* repressors Spo0A at the *abrB* promoter and MerR at the *merT* promoter (in the absence of mercury) have been shown to impair the transition from closed to open complexes. Among the second group, the GalR repressor has to be mentioned (see below for a detailed description).

Repressors Inhibiting Promoter Clearance Clearance of RNA polymerase from the promoter is not straightforward. After the formation of an initiated complex, RNA polymerase must break contacts with the promoter and with any transcriptional regulator that could be in contact with it. This can pose some problems. Indeed, promoter strength is known to be a compromise between efficient recognition by RNA polymerase (synonymous of tight binding) and efficient clearance, which is required to disrupt that binding. For example, RNA polymerase can be stalled at the +6 to +12 region *in vivo* when its binding to consensus promoter elements is too tight. Therefore, a promoter can be repressed by interfering with the clearance process; and bacteria have exploited this possibility. An interesting example is that of the H-NS protein at the *rrnB* P1 promoter. In this case, H-NS binds to a region that overlaps promoter sequences, changing its structure, but does not inhibit subsequent binding of RNA polymerase. Although open complexes are efficiently formed, the generation of initial transcripts larger than three nucleotides in length is inhibited. It was proposed that repression occurs because H-NS changes the conformation of the open complexes.

Repression by Looping Multiple repressor molecules bind to promoter-distal sites; and repression is caused by DNA looping. Under these circumstances, the RNA polymerase is unable to bind to its promoter, preventing transcription initiation in the looped region. This mechanism is exemplified by the GalR repressor.

Repression by Sequestration of an Activator Here, the repressor functions as an anti-activator by binding to an activator protein. The best studied examples are CytR-repressed promoters. Repression by the CytR repressor occurs through direct interactions between CytR and CAP, where both proteins bind to the DNA upstream of the promoter.

F. Rojo **1999**, Repression of transcription initiation in bacteria, *J. Bacteriol.* 181, 2987–2991.
F. Rojo **2001** Mechanism of transcriptional repression, *Curr. Opin. Microbiol.* 4, 145–151.

6.2.8.1.2 Modulation of Repressor Activity

Transcription factors couple the expression of genes to external (environmental) or internal signals. It follows that their activity has to be modulated depending on

the presence or absence of the appropriate signal. Different mechanisms are used to achieve this aim:

- modulation by small ligands which either signal the presence of a nutrient (e.g., the LacI repressor) or the absence of a metabolite (e.g., the TrpR repressor)
- modulation by temperature (e.g., the RheA repressor)
- modulation by proteolysis (e.g., the LexA repressor)
- modulation by a corepressor (e.g., the HspR repressor)
- modulation by the GroE chaperonin team (e.g., the HrcA repressor)
- modulation by covalent modification of the regulator through phosphorylation (e.g., the response regulators)
- modulation by DNA looping (e.g., the GalR repressor)
- the LacI repressor senses the presence of the nutrient lactose.

Regulation of the *lac* operon serves as the paradigm to develop the concept of the operon model by Jacob and Monod in 1961. Transcription of this operon is initiated from the *lac* promoter resulting in the expression of the three structural genes *lacZ*, *lacY* and *lacA* (Fig. 6.16). These genes code for β-galactosidase, lactose permease and transacetylase. The first enzyme, β-galactosidase, has two activities, a major and a minor one. The major activity is to cleave lactose into glucose and galactose; and the minor activity is used to convert a few molecules of lactose to allolactose, the true inducer of the *lac* operon. The lactose permease is an integral inner membrane protein responsible for taking up lactose, concomitant with one proton per sugar molecule (symporter). The function of the third protein, the transacetylase, in unknown. In the absence of lactose, the operon is expressed at a low level, representing the default state. This is ensured by the LacI repressor when binding to the *lac* operators O_1, O_2 and O_3. While the O_1 operator overlaps promoter sequences, operators O_2 and O_3 are located 401 bp downstream and

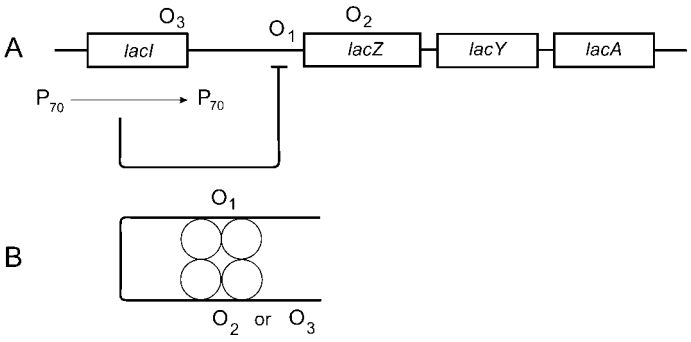


Fig. 6.16 Regulation of the *lac* operon by the LacI repressor. (A) The *lacZYA* operon is under the negative control of the LacI repressor binding first to the *lac* operator O_1 located immediately downstream of the *lac* promoter and then to either O_2 (located within the coding region of *lacZ*) or O_3 (located within the coding region of *lacI*) forming a DNA loop (B).

92 bp upstream, respectively, from the O_1 operator. O_1 alone represses transcription about 20-fold. LacI is a stable tetramer that can simultaneously bind to two operators, O_1 and either O_2 or O_3, generating a DNA loop that modifies promoter geometry, increasing repression by an extra 50-fold (Fig. 6.16).

When lactose is added to growing cells, it diffuses through outer membrane porins into the periplasm and binds to the LacY permease present in a few copies. After uptake into the cytoplasm, the residual β-galactosidase activity present in the cell converts a few molecules of lactose into allolactose; and these allolactose molecules bind (in a 1:1 stoichiometry) to the LacI repressor, causing a conformational change and subsequent dissociation from the *lac* operators, resulting in induction of the *lac* operon. But the operon is also under positive control by the CRP (CAP) protein (see below). If both glucose and lactose are present in the medium, the *lac* operon only becomes induced if cells have consumed the glucose. This phenomenon is designated catabolite repression; and it ensures that bacteria first catabolize high-energy sources such as glucose.

The TrpR Repressor Senses the Absence of Tryptophan

The five genes of the biosynthetic *trp* operon of *E. coli* are responsible for the synthesis of the amino acid L-tryptophan. This operon is under the negative control of the 12.2-kDa TrpR repressor protein which binds to its operator located downstream of the P_{trp} promoter. In addition, the TrpR repressor binds to two additional operators, one upstream of the gene *trpR* coding for the repressor (autoregulation) and one upstream the *aroH* operon involved in the synthesis of chorismate. But the TrpR repressor can bind only to the three operators when a certain level of the corepressor tryptophan is present in the medium. Tryptophan binds to the TrpR aporepressor and changes its conformation in such a way that it can bind to its operators. If the tryptophan concentration within the cell falls below a threshhold value, free aporepressor is unable to exert its function and the three operons are transcribed. Therefore, the TrpR repressor senses the absence of tryptophan.

The RheA Repressor is Modulated by Temperature In *S. albus*, a small (18-kDa) heat-shock protein is strongly induced after a sudden increase in temperature. The gene *hsp18* coding for this protein is under negative control by the transcriptional repressor RheA encoded by the *rheA* gene located 150 bp upstream from and in the opposite orientation to *hsp18*, whereby both genes form a *divergon*. Divergons are transcriptional units of opposite polarity separated by a common regulatory region containing two promoters and sometimes, in addition, binding sites for a transcriptional regulator, as in the present case (see Section 6.2.8.4 for detailed description). While no *hsp18* transcript can be detected at 30 °C, high concentrations are present at 41 °C. A detailed analysis of the purified RheA repressor revealed a reversible temperature-related change in RheA conformation, reflecting a transition between an inactive and an active form. While RheA binds its operators at low temperature, 30 °C or below, preventing expression of *hsp18* and autoregulating its own expression, it becomes immediately inactivated in response to a heat shock.

The LexA Repressor is Modulated by Proteolysis LexA of *E. coli* (and many other bacterial species) is a negative regulator of the SOS regulon which is induced upon severe damage of DNA (see Section 5.3.9). The LexA protein consists of two separable domains: a dimerization domain responsible for dimerization and a DNA binding domain involved in binding to the operators, called SOS boxes. DNA damage leads to the formation of single-stranded DNA, which in turn converts RecA into an activated form, termed RecA*, able to promote autocleavage of target proteins such as LexA. Binding of RecA* to LexA induces self-cleavage of LexA between Ala111 and Gly112, resulting in two fragments of similar size. When the DNA damage has been repaired by one of several mechanisms, the RecA* protein disappears, leading to the reappearance of intact LexA due to *de novo* synthesis, resulting in the turn-off of the SOS regulon. Here, the activity of the repressor is modulated by proteolysis.

The HspR Repressor is Modulated by the Corepressor DnaK The *dnaK* operon of *Streptomyces coelicolor* codes for the DnaK chaperone system and HspR, the transcriptional repressor of the operon. HspR confers repression of the operon by binding to several inverted repeats in the promoter region. HspR by itself is unable to bind to its operator; and it needs DnaK as a corepressor. In the absence of heat shock, both form a stable complex at the operator to reduce transcription of the *dnaK* operon. After a heat shock, the sudden appearance of nonnative proteins titrate DnaK, thereby resulting in the induction of the *dnaK* operon. When the nonnative proteins have been removed from the cell, free DnaK forms a complex with HspR to confer repression of the *dnaK* operon again.

The HrcA Repressor is Modulated by the Molecular Chaperone GroEL The HrcA repressor is present in more than 100 bacterial species and represents the most widespread regulatory system in eubacteria. In all these species, it regulates at least the *groESL* operon; and in many it regulates additional heat shock genes such as the *dnaK* operon (see Section 9.2). The HrcA protein, binding as a dimer to its operator, in most cases a perfect inverted repeat of 9 bp, is present in two conformations, one active and the other inactive. The conversion from the inactive to the active conformation able to interact with its operator is promoted by the GroE chaperonin team. It ensures that enough active HrcA dimer is present within the cells to reduce expression of the appropriate heat-shock operons. After a sudden increase in temperature, the GroE chaperonin machines are titrated by the nonnative proteins, as mentioned before for the DnaK team. This leads to an increase in the amount of inactive HrcA followed by an enhanced expression of those operons controlled by HrcA. The more nonnative proteins are removed from the cytoplasm, the more GroEL becomes available to promote conversion of the inactive HrcA molecules into the active form and repression is resumed.

Repressors Belonging to the Response Regulator Family are Modulated by Covalent Modification Response regulators are members of the two-component signal transduction system family (see below, under Section 6.2.8.5). In short, these sys-

tems consist of two members, a sensor kinase and a response regulator. Upon sensing a signal, the sensor kinase undergoes autophosporylation and then transfers the phosphate to the response regulator. With a few exceptions, all response regulators are DNA-binding proteins acting as repressors, activators or dual regulators. In the absence of the appropriate environmental signal, these transcriptional regulators are largely inactive, while phosphorylation at an invariant aspartate residue converts them into a DNA-binding protein. Inactivation of the phosphorylated response regulator occurs either passively by *de novo* synthesis of new molecules which are not any more phosphorylated or actively by a phosphatase removing the phosphate.

The GalR Repressor Modulates Repression by DNA Looping GalR represses two promoters of the *gal* operon, P_1 and P_2, located 5 bp apart, by binding to two operators named O_E and O_I. Operator O_E is located at –60.5, while O_I is at +53.5. In the presence of the chromatin-associated protein HU, GalR binds to both operators and generates a DNA loop that inhibits transcription from the P_1 and P_2 promoters (Fig. 6.17). The HU-GalR-DNA loop complex had been termed the *repressosome*. First, the GalR 37-kDa protein forms a dimer to which Hu can bind. Next, this complex interacts with the two operators and the two GalR dimers contact each other, forming a tetramer. HU spontaneously dissociates from GalR and binds to the bend without sequence specificity, leading to stabilization of the structure. When the loop is not formed because HU is absent or because only the O_E operator is present, GalR does not inhibit RNA polymerase binding to the promoters, although it represses P_1 (not P_2) by impairing the transition from closed to open complex formation.

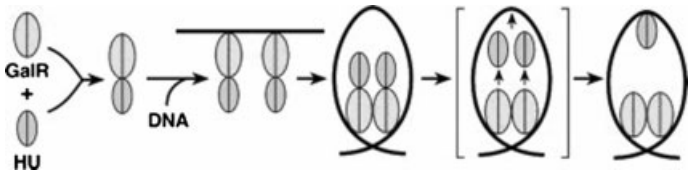


Fig. 6.17 Pathway of repressosome formation. GalR dimers recruit HU and bind to the two operators. Two GalR dimers interact transiently forming a tetramer and generate a DNA loop. HU will spontaneously dissociate, bind to the bend stabilizing the structure. S. Roy **2005**, *Biochemistry* 44, 5373–5380; Fig. 9.

6.2.8.2 Transcriptional Activators

Activation of transcription is a common way to regulate gene expression in both prokaryotes and eukaryotes. In bacterial cells, three different mechanisms of transcription activation have been described:

- recruitment of the RNA polymerase
- prerecruitment of the RNA polymerase
- promoting isomerization of the RNA polymerase holoenzyme containing σ^{54}.

6.2.8.2.1 Recruitment of the RNA Polymerase by Activators

Recruitment Recruitment is commonly accepted as the main mechanism of transcription activation in bacteria. In the typical pathway of recruitment, the activator first binds its DNA target. Then, through protein–protein interactions with the DNA-bound activator, RNA polymerase is recruited to the promoter and a stable, open complex is formed. At these promoters, the polymerase cannot occupy the DNA fully, since one or more of the DNA sequence elements, the –35 and the –10 regions, is defective. Defects in these elements, more often in the –35 sequence, provide evidence for a physiological basis for an activator function. The best studied transcription activator is the CAP (CRP) protein. While most recruitments just involve one activator, there are also examples that more than one protein is involved, called transcription at complex promoters.

Transcription Activation at Simple Promoters At many promoters, activation occurs in a simple way involving a single transcriptional activator. Simple activation occurs by one of three different mechanisms:

- class I activation
- class II activation
- activation by a conformational change.

Class I Activation The activator protein binds to a target site located upstream of the –35 region of the promoter and recruits the RNA polymerase by interacting with the αCTD. Since the linker joining the αCTD and the αNTD of the α subunits is flexible, the activator can bind at several locations upstream of the promoter. The best example is CAP (for *c*atabolite *a*ctivator *p*rotein; also known as the *c*AMP *r*eceptor *p*rotein, CRP) at the *lac* promoter.

Class II Activation The activator protein binds to a target overlapping the –35 region of the promoter and recruits the RNA polymerase by making contact with domain 4 of the σ subunit. This mechanism allows very little flexibility in the positioning of the transcriptional activator. Though the binding site remains unchanged, the activator can also contact other subunits of the RNA polymerase, e.g., the αNTD. Here, a well studied example is the activation of the phage λ PRM (for *p*romoter of *r*epressor *m*aintenance) by the CI protein. This promoter is active as part of a λ prophage and guarantees sufficient amounts of λ repressor by autoregulation. CAP activates transcription at more than 100 promoters and was the first activator to be purified and have its three-dimensional structure determined. This activator functions by binding, in the presence of the allosteric effector cAMP, to specific DNA sequences in or near target promoters and thereby enhances the ability of the polymerase holoenzyme to bind and initiate transcription. CAP has a molecular mass of 45 kDa and acts as a homodimer. Each subunit consists of two domains: the N-terminal domain is responsible for dimerization and for the interaction with the allosteric effector cAMP, while the C-terminal domain is responsible for binding to the DNA through a HTH motif.

Activation by Conformation Change Here, the activator alters the conformation of the target promoter to allow binding of the RNA polymerase with its –35 and/or –10 regions. This requires that the activator binds at or very near the promoter regions. In the case of the MerR-type activators, they bind to the spacer region of the promoter to twist the DNA in such a way that the –35 and –10 regions are reoriented to allow binding of the RNA polymerase.

Transcription Activation at Complex Promoters While a single activator–RNA polymerase contact is sufficient at simple promoters, many naturally occurring activator-dependent promoters are complex and need contact with two or even more different proteins. They are co-regulated either by a repressor or by a second activator (or by both). Regulation by more than one transcriptional regulator allows the cell to respond to different environmental signals. We will discuss different complex promoters:

- repositioning of a primary by a secondary activator (e.g., MalT by CAP)
- co-dependence on independent contacts by two activators (e.g., Fis and CAP)
- co-dependence due to cooperative binding of activators (e.g., MelR and CAP)
- co-dependence due to bacterial nucleotide proteins (e.g., Fis, IHF and NarP/NarL)
- modulation by an epigenetic mechanism (e.g., Dam methylase)
- anti-activators (e.g., TraM-TraR, PspA-PspF and NIFA-NIFL)
- subcellular relocalization of the activator.

Repositioning of a Primary by a Secondary Activator The *E. coli* maltose system consists of ten genes encoding proteins dedicated to the uptake and metabolism of maltose and maltodextrins. The *malEFG* gene products, together with the *malK* gene product, encode the specific maltose/maltodextrin binding–protein-dependent ABC transporter. These genes are under the control of MalT, a specific transcriptional activator of 901 amino acids. MalT belongs to a class of bacterial transactivators, the MalT or LAL family. The regulatory region of the *malK* promoter comprises two series of MalT-binding sites, one with high affinity, another with low affinity, separated by three CAP-binding sites. Occupation of the higher set of sites occurs in the absence of CAP and does not lead to *malK* activation. In the presence of CAP, however, the set of sites with low affinity is properly positioned with respect to the promoter. The synergistic action of MalT and CAP relies on MalT repositioning via the formation of a nucleoprotein structure.

Co-dependence on Independent Contacts by Two Activators The *E. coli proP* P2 promoter controls expression of an integral membrane transporter of proline, glycine betaine and other osmoprotecting compounds. Transcription from this promoter is dependent on the Fis protein as well as the stationary sigma factor, σ^{38} (σ^{S}). Fis activates *proP* P2 transcription when bound to a site centered at –41, where it overlaps the –35 binding element for the sigma subunit of RNA polymerase. The *proP* regulatory region also contains a binding site with high affinity for the cAMP receptor protein (CRP = CAP) located at –121.5 relative to the P2 pro-

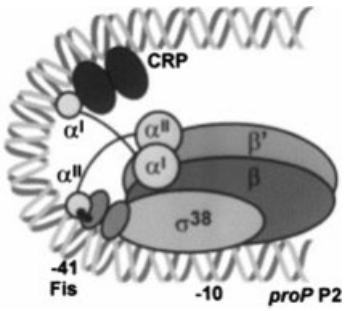


Fig. 6.18 Architecture at the *E. coli proP* P_2 promoter. Activation of the promoter needs binding of the two proteins Fis and CRP (CAP), where both protein dimers are contacted by one αCTP each. S.M. McLeod, et al. **2002**, *J. Mol. Biol.* 316, 517–529; Fig. 7.

moter start site. In the absence of Fis, CRP-cAMP provides little activation. However, Fis and CRP-cAMP together activate transcription synergistically. The architecture of the activated transcription complexes at the *proP* P2 and *rrn* P1 promoters is shown in Fig. 6.18.

Co-dependence Due to Cooperative Binding of Activators MelR is a melibiose-triggered transcription activator that belongs to the AraC family of transcription factors which is involved in the activation of the *melAB* operon. While *melA* codes for an α-galactosidase, *melB* encodes a thiomethylgalactosidase permease II. In the absence of melibiose, it binds to three different sites centered at positions –100.5 (site 1), –120.5 (site 1′) and –62.5 (site 2; Fig. 6.19). CAP is unable to bind upstream of the *merAB* promoter in the absence of MelR. But if MelR has bound to these three sites, it interacts at a site located between them. Thus, binding of CAP strictly depends on previous binding of MelR. But binding these two proteins together is not enough to activate transcription of the *melAB* operon. Upon the addition of melibiose, the sugar binds to MelR causing a conformational change and allowing MelR to occupy a fourth site called 2′. Now, the whole protein complex is able to activate the *melAB* operon (Fig. 6.19).

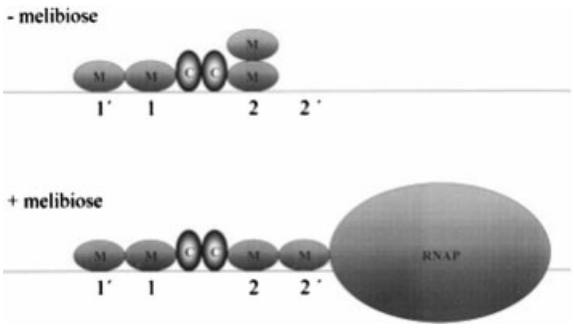


Fig. 6.19 Activation of the *E. coli melAB* operon is dependent on two different activators. In the absence of melibiose, the MelR protein (designated as ''M'') binds to three different sites designated as 1′, 1 and 2. They sandwich a binding site for CAP (C). If melibiose is present, it binds to MelR causing a conformational change allowing MelR to occupy site 2′. T.A. Belyaeva, et al. **2002**, *Mol. Microbiol.* 36, 211–222; Fig. 8.

Co-dependence Due to Bacterial Nucleoid Proteins In *E. coli*, the expression of several genes required for anaerobic respiration is controlled by the FNR (for *f*umarate and *n*itrate *r*eduction) protein, an anaerobically induced transcription activator. In many cases, FNR binds as a dimer to a 22-bp DNA sequence centered near position –41, either to directly activate transcription initiation or to be dependent on additional activators, such as NarL or NarP (Fig. 6.20). These activator proteins are regulated by the presence of nitrite (NarL) or nitrate (NarP) ions in the medium. Both activators belong to two different two-component signal transduction systems where the sensor kinase NarX phosphorylates NarL and NarQ NarP. The *E. coli nir* operon codes for a cytoplasmic NADH-dependent nitrite reductase, which reduces nitrite ions to ammonia. Transcription of this operon starts at a single promoter and is activated by two environmental signals: the absence of oxygen and the presence of nitrite or nitrate ions in the growth medium. Upstream of the *nir* promoter, an FNR binding site occurs at –42 and a NarL/NarP site is centered at position –69.5. Two additional proteins bind upstream of the promoter, Fis at –142 and at –23 and IHF at –88 (Fig. 6.20). Both nucleoid-associated proteins organize the promoter into an ordered nucleoprotein repression complex; and this complex is disrupted by NarL~P or NarP~P.

Modulation by an Epigenetic Mechanism Another example of a complex activation mechanism has already been described in detail (see Section 6.1.2) and involves the *pap* promoter. Methylation at two GATC sequences creates an epigenetic switch that is flipped under the influence of PapB and PapI and triggered by the appearance of hemi-methylated DNA.

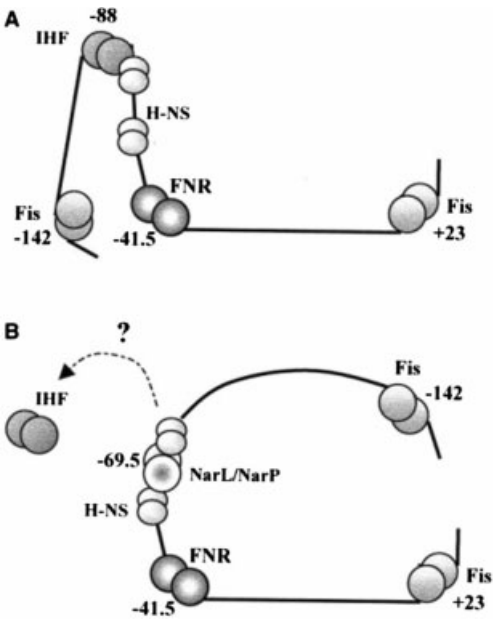


Fig. 6.20 Activation of the *E. coli nir* promoter. (A) Binding of the nucleoid-associated proteins Fis and IHF converts the promoter region into a nucleo-protein structure preventing FNR to activate transcription. (B) Phosphorylation of NarL or NarP due to the presence of nitrite or nitrate, respectively, allows their binding to the promoter region thereby destroying the repression complex and allowing activation of the *nir* promoter. D.F. Browning, et al. **2000**, *Mol. Microbiol.* 37, 1258–1269.

Anti-activators Conjugative transfer of the Ti plasmid of *A. tumefaciens* is controlled in a population density-dependent manner. In this system, TraR activates transcription of target operons in response to the accumulation of its cognate autoinducer, a *h*omoserine *l*actose (HSL) derivate produced by TraI, an acyl-HSL synthase (see Section 6.2.10.1). TraR, in dimer form, binds to 18-bp inverted repeat elements termed *tra* boxes, located in the promoter regions of the *tra* and *trb* operons, leading to the activation of the *tra* genes responsible for Ti plasmid conjugation. The TraR-mediated *tra* gene expression is modulated by the TraM protein. TraM inhibits transcriptional activation of the *tra* regulon by directly interacting with TraR. Thus, TraM acts an anti-activator and targets the C-terminal region of TraR, including the HTH domain. The 11.2-kDa TraM proteins consists of 101 amino acids, forms a homodimer where regions in the middle of the protein are involved in dimerization and regions at both ends of TraM (ca. residues 20–50 and 67–96) are required for interaction with TraR.

A second anti-activator is PspA, which sequesters the activator protein PspF. The *pspA* (for *p*hage *s*hock *p*rotein *A*) gene is the first gene of a tetracistronic operon, where *pspBCD* all code for inner membrane proteins. This operon is induced by multiple stress factors: filamentous phage infection, mislocation of some envelope proteins (e.g., secretin), extremes of temperature, osmolarity or ethanol concentration and the presence of proton ionophores such as carbonylcyanide *m*-chlorophenylhydrazone (CCCP). A unifying consequence of these conditions may be dissipation of the *p*roton-*m*otive *f*orce (PMF), though this has not been tested in most cases. A working model for the Psp response is presented in Fig. 6.21. In the uninduced state, the transcriptional activator PspF is kept inactive by interaction with PspA. Induction of the Psp response, for example by mislocation of secretin (an outer membrane pore-forming protein) and/or dissipation of the PMF, is sensed by PspB and PspC, leading to the sequestration of PspA. This in turn leads to the release of PspF which activates transcription of the *psp* operon, its own gene and the unlinked *pspG* gene. Both *pspD* and *pspG* are not part of the signal transduction pathway and their function remains elusive.

The third example concerns the NifA-NifL pair of *A. vinelandii*, which controls transcription of nitrogen fixation genes in response to the concentration of fixed nitrogen and extracellular oxygen. NifA is a multidomain σ^N-dependent activator which consists of an N-terminal domain of unknown function, a conserved central domain that catalyzes ATP hydrolysis and most probably interacts with the $E\sigma^N$ holoenzyme and a C-terminal domain containing a HTH motif involved in recognition of the upstream activator sequences. The anti-activator NifA inhibits NifL via a concerted mechanism in which hydrolysis of ATP, DNA binding and, potentially, interaction with $E\sigma^N$ are controlled to prevent transcription activation under detrimental environmental conditions. NifL senses the redox state within the cytoplasm *via* an FAD cofactor located within the N-terminal PAS domain. Deletion of this domain or removal of the FAD moiety renders NifL insensitive to the redox state. The NifL C-terminal domain binds adenosine nucleotides in a Mg^{2+}-dependent manner with a ten-fold higher affinity for ADP than ATP, leading to a conformational change. Both ADP and ATP stimulate the inhibitory activ-

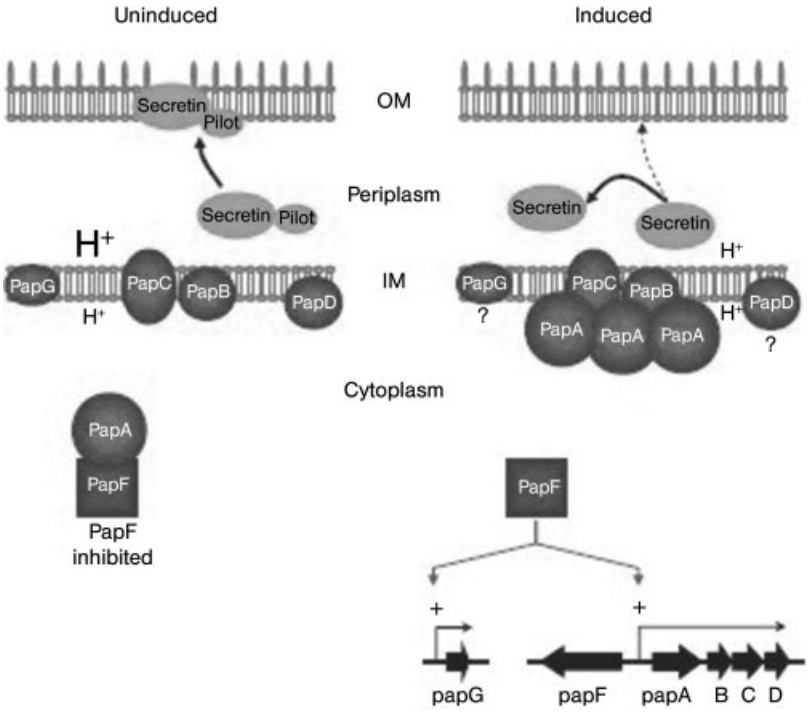


Fig. 6.21 Working model for the release of the PspF activator protein from its anti-activator PspA. In the uninduced state, PspA and PspF form a stable complex. In the induced state, e.g., by mislocation of secretin, the inner membrane protein PspB and/or PspC may sense an inducing signal followed by binding of PspA thereby distracting it from PspF. The free transcriptional activator PspF will activate the two operons shown. A.J. Darwin **2005**, *Mol. Microbiol.* 57, 621–628; Fig. 1.

ity of NifL by increasing the stability of the NifA-NifL interaction. In the presence of ADP, NifL is able to interact with both the N-terminal and AAA^+ domains of NifA. Inhibition by the ADP-bound form of NifL is antagonized by the binding of the 2-oxoglutarate to the N-terminal GAF domain of NifA, enabling the NifL-NifA system to respond to a key metabolic signal of the carbon status.

Subcellular Relocalization of the Activator Transcriptional activators can be sequestered by cognate transmembraneous transporter proteins. Binding or release of the activator protein is directly controlled by the activity of the transporter. According to this principle, the bacterial cell is able to measure the external substrate concentration. One example is the *E. coli* ABC transporter responsible for maltose uptake (the $MalEFGK_2$ complex), one of the best studied of the binding protein-dependent multicomponent systems. MalE (or MBP, for *m*altose *b*inding *p*rotein) is a soluble high-affinity maltose–maltodextrin binding protein (dissociation constant, K_d = 1 μM) located in the periplasm and the substrate-recognition site of

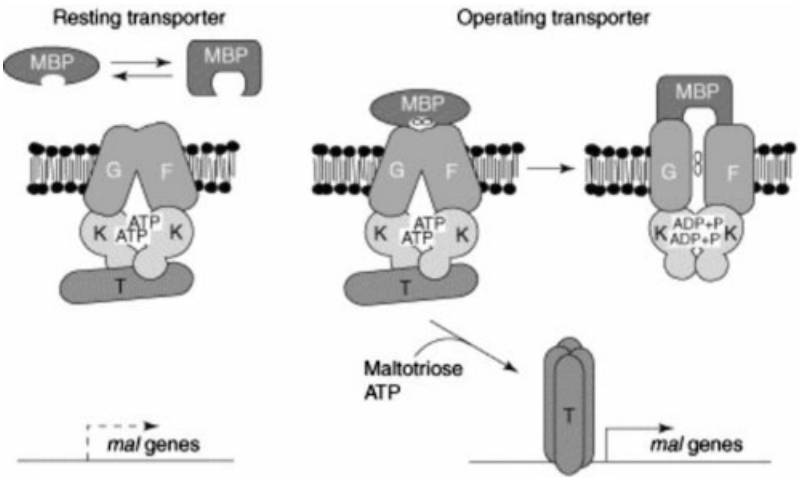


Fig. 6.22 The maltose ABC transporter. A. Böhm, W. Boos **2004**, *Curr. Opin. Microbiol.* 7, 151–156; Fig. 1.

the system (Fig. 6.22). The two integral membrane proteins MalF and MalG form a transport channel through the inner membrane and the MalK homodimer is positioned in the cytoplasm and associated tightly with the MalF-MalG complex. Hydrolysis of ATP drives active transport against a steep concentration gradient and is assumed to deliver protein conformational energy to MalF-MalG in the transport process. Binding of substrate-loaded MalE to the MalF-MalG complex on the periplasmic side triggers a chain of conformational changes, leading to ATP hydrolysis by MalK and resulting in substrate release into the cytoplasm. But MalK exerts another function as a negative regulator of MalT, the transcriptional activator of the maltose regulon. The presence of the two effector molecules ATP and maltotriose induce dimerization of MalT, followed by cooperative binding to an array of sites located upstream of the –35 promoter elements. The unliganded form of MalT is monomeric and bound by MalK (see Fig. 6.22) and in addition by MalY and Aes.

6.2.8.2.2 **Pre-recruitment of the RNA Polymerase by Activators**

Pre-recruitment was discovered quite recently and is defined as binding of the RNA polymerase holoenzyme by the activator protein in the cytoplasm, followed by scanning of the DNA by the binary complex for activator-dependent promoters. This new mechanism has been elucidated for the SoxS activator. But circumstantial evidence suggests that additional activators including MarA, Rob and TetD act by pre-recruitment.

The SoxS Activator The SoxRS system of *Escherichia coli* is responsible for the cell's defense against superoxide anion, nitric oxide and redox-cycling com-

pounds, like paraquat, that endogenously generate superoxide. The regulatory process leading to induction of the regulon during oxidative stress is unusual in that it occurs in two temporal stages of transcription, the first mediated by SoxR and the second by SoxS (see Section 9.4.2). In short, SoxR is expressed constitutively in a form able to bind its sole DNA target, a site in the promoter of the *soxS* gene, but unable to activate *soxS* transcription. However, oxidation of the 2Fe–2S centers of SoxR induces a conformational change in the protein that allows it to activate *soxS* transcription. Then, SoxS, synthesized *de novo* in response to the oxidative stress activates the transcription of the ~40 genes of the regulon. Thus, SoxR is the sensor and transmitter of oxidative stress, while SoxS is the response regulator. Another property distinguishing SoxS from most other bacterial transcription activators is that its DNA binding site is highly degenerate. Information content analysis of the SoxS binding sites and direct analysis by gel shift assay has indicated that fast-growing cells of *E. coli* contain ~65 000 SoxS binding sites, called soxbox. Since the SoxS regulon contains only ~40 genes, it is unclear how SoxS is able to mount the defense response against oxidative stress when it is synthesized *de novo* and the total number of binding sites per cell far exceeds the number of functional binding sites residing in SoxS-dependent promoters. However, the maximum number of SoxS molecules per cell has been determined to be about 2500. This 26-fold disparity between the total number of SoxS binding sites per cell and the maximum number of SoxS molecules per cell indicates that transcription activation by SoxS is not likely to occur by the usual recruitment mechanism. To comply with this problem, pre-recruitment has been suggested as the mechanism to distinguish SoxS binding sites residing at the proper position and orientation in SoxS-dependent promoters from the vast excess of sequence equivalent. In pre-recruitment, *de novo* synthesized SoxS forms a binary complex with RNA polymerase in the cytoplasm, thus acting as a co-sigma factor. The SoxS-RNA polymerase binary complex then scans the chromosome for SoxS-dependent promoters, using the σ^{70} factor of the polymerase to search for the –10 and –35 promoter recognition elements and SoxS to search for soxboxes located in the proper position and orientation for binding and transcription activation. SoxS interacts with αCTD sequences involved in binding to the UP element (Fig. 6.23). By doing so, SoxS also distracts RNA polymerase from the strong promoters preceding the rRNA operons.

Other Activators Acting by Pre-recruitment MarA and Rob form together with SoxS a closely related subset of the AraC/XylS family of activators, sharing ~42% amino acid sequence identity over the length of SoxS. All three proteins bind DNA as monomers, have a relatively weak affinity for DNA (K_d = ~10^{-8} M) and recognize the same highly degenerate 20-bp binding sites located either upstream or overlapping the –35 element. Although all three activator proteins recognize the same degenerate DNA binding site, they activate transcription of regulon genes to different degrees and in response to different signals. While transcription of *soxS* is activated in response to changes in the redox potential of the cell caused by the action of redox-cycling compounds, *marA* transcription is induced

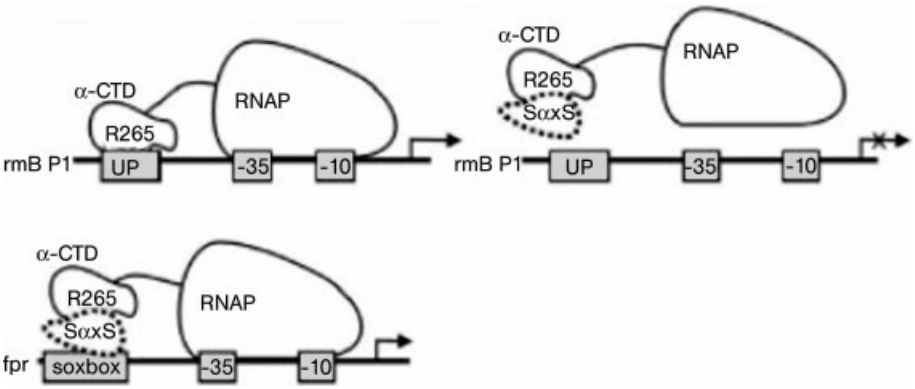


Fig. 6.23 Transcription activation by pre-recruitment of the RNA polymerase by SoxS. RNA polymerase binds to the promoters of the *rrn* operons with high efficiency through interaction between the αCTD and the UP element (left part). SoxS binds to the αCTD preventing its interaction with the UP element (middle part). The thus pre-recruited RNA polymerase is distracted to promoters with an upstream soxbox element. I.M. Shah, R.E. Wolf **2004**, *J. Mol. Biol.* 343, 513–532; Fig. 5.

by antibiotics and weak aromatic acids. Rob, in contrast, is expressed constitutively at 5000–10000 molecules per cell, but these molecules appear inaccessible to the RNA polymerase, as they are present in discrete immunostaining foci. Treatment of cells with dipyridyl or bile salts increases the Rob activation of target genes. TetD is another member of the AraC/XylS family of transcription activators. The *tetD* gene resides in the transposon Tn*10* and neither its inducer nor its function is known. It has been suggested that MarA, Rob and TetD function as a co-sigma factor and activate transcription by the pre-recruitment mechanism. MarA has been shown to form a binary complex with RNA polymerase *in vitro*.

6.2.8.2.3 Isomerization of the RNA Polymerase by Activators

The third mechanism is specific for $E\sigma^{54}$. While the holoenzyme binds to σ^{54}-dependent promoters as a closed complex, it is unable to carry out the *isomerization step* to the open complex. This isomerization step is promoted by interaction with a transcriptional activator. Here, NifC has been studied in great detail. To date, over 600 activators have been defined by sequence alignment.

Activation of σ^{54}-dependent Promoters

As already mentioned, $E\sigma^{54}$ binds to its promoter but is unable to carry out the isomerization step (resulting in strand separation) by its own. This step requires a transcriptional activator that binds to sites that are effective regardless of distance and orientation. In this respect, the activator binding sites are analagous to eukaryotic *enhancers*; and the activators are often called enhancer binding proteins. Enhancers by definition still function when they are moved far away from the promoter. Other activator binding sites function poorly when they are moved away

from their promoter; and such a site is designated *UAS* for *u*pstream *a*ctivating *s*equence. It should be mentioned that the UAS is not essential for activation in many cases if the activators are expressed at higher levels than normal. The activators interact with $E\sigma^{54}$ from these binding sites; and this interaction sometimes requires a DNA-bending protein when the enhancer or UAS is too far away from the promoter to allow direct contact between the bound activator and the RNA polymerase. In *E. coli*, the IHF and the ArgR (*ar*ginine *r*epressor) have been identified as DNA bending proteins. When transcription from a σ^{54}-dependent promoter does not require a DNA bending protein, DNA curvature facilitates the interaction between the activator and RNA polymerase. It should be mentioned that the DNA binding site becomes dispensible when the activator protein is expressed at higher than normal levels. Under these circumstances, the activator may contact the closed complex directly from solution to activate transcription. The σ^{54}-dependent response regulators generally consist of three domains, an N-terminal regulatory domain, the AAA^+ central domain where a conserved GAFTGA loop is involved in activation, and a C-terminal DNA-binding domain. The linkers between domains are variable in length and diverse in sequence. The central domains are well conserved among σ^{54}-dependent activators and can be subdivided into seven well conserved regions designated C1 to C7. Functions have been proposed for some regions, based on amino acid sequence comparisons. Region C1 forms a structural motif known as a phosphate loop or Walker type A sequence that is common to a variety of nucleotide binding proteins. The C_4 region is similar to a motif known as a Walker type B sequence that is also common to many nucleotide binding proteins. Region C3 is thought to participate in productive interactions between the activator and $E\sigma^{54}$. The inactivated proteins are dimers, whilst formation of higher oligomers, coupled with ATP hydrolysis, is apparently necessary to activate transcription by the σ^{54}-RNA polymerase complex. The following mechanisms have been identified controling the interaction of the regulatory domain with the ATP binding domain:

- phosphorylation within the N-terminal receiver domain causing oligomerization (e.g., NtrC)
- phosphorylation within the receiver domain relieves inhibition imposed by the nonphosphorylated receiver domain upon the central AAA^+ domain (e.g., DctD)
- binding of a small ligand to the N-terminal domain (e.g., xylene to XylR)
- use of a second partner protein which imposes negative control (e.g., N-terminal domain of NifA, targeted by NifL, which causes inhibition of NifA activity).

This last inhibition occurs partly by diminishing NifA ATPase activity and the ability of NifA to bind UAS DNA (see above, under anti-activators).

The NtrC Activator One of the most extensively studied regulators is the general *nit*rogen *r*egulatory protein *C* (NtrC) that controls the nitrogen assimilation in many bacteria. NtrC is member of a two-component signal transduction system

(see below), where the NtrB protein functions as the sensor kinase. NtrB senses a drop in the intracellular concentration of glutamine, leading to autophosphorylation and transfer of the phosphoryl group to the NtrC response regulator. Phosphorylation causes NtrC~P to form homotetramers and higher oligomers, with subsequent binding to the enhancer-like NtrC binding sites. This leads to the expression of several genes, among them *glnA* coding for a glutamine synthetase catalyzing the synthesis of glutamine from glutamate and ammonia.

The DctD Activator The C_4-*di*carboxylic acid *t*ransport protein *D* DctD of *S. meliloti* is an activator of $E\sigma^{54}$ that activates transcription from *dctA*, which encodes a permease for C_4-dicarboxylates and whose expression is inducible with C_4-dicarboxylates. DctD contains three functional domains, an N-terminal regulatory domain, a central domain responsible for transcriptional activation and ATP hydrolysis, and a C-terminal domain with a HTH motif responsible for recognition and binding to the UAS located upstream of the *dctA* promoter region. The regulatory domain inhibits the transcriptional activation and ATPase activities of the central domain. Region C3 has been shown to contact both σ^{54} and the β subunit. DctD is part of a two-component signal transcduction system together with DctB, where *dctD* codes for the response regulator and *dctB* for the sensor kinase. In free-living conditions, *dctB* and *dctD* are constitutively expressed at a low level. When C_4-dicarboxylates are present in the periplasm, DctB phosphorylates DctD which then binds specifically and cooperatively to a tandem UAS located –94 to –154 upstream of the *dctA* transcriptional start site, followed by activation of *dctA* through catalyzing the isomerization step.

The XylR Activator Strains of *Pseudomonas* are known for their ability to degrade a wide variety of aromatic compounds. This ability is usually determined by large catabolic operons often present on plasmids, such as pWW0. Strains carrying pWW0 are able to grow on toluene, *m*-xylene and *p*-xylene as the only carbon source, which are transformed to pyruvate and acetyl-CoA. Expression of the TOL operon involved in bioconversion of these compounds is driven by the σ^{70}-dependent promoter *Pu* (Fig. 6.24). In the presence of a suitable aromatic effector (e.g., *p*-xylene), this promoter is activated at a distance by the XylR activator protein with the assistance of IHF facilitating the appropriate geometry at *Pu* and helping to recruit $E\sigma^{54}$. XylR-mediated transcription activation stands as a model for a larger group of regulatory proteins for different aromatic pathways, such as DmpR (degradation of phenol), TouR (degradation of toluene and *o*-xylene) and HbpR

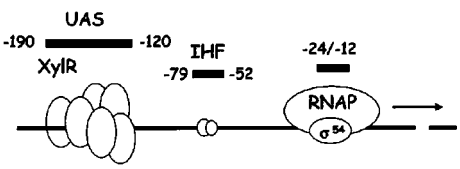


Fig. 6.24 Organization of the *Pu* promoter region of pWW0. This schematic drawing exhibits the $E\sigma^{54}$ at its promoter, the location of the binding site for IHF and the UAS to which XylR binds. M. Carmona, et al. **2005**, *J. Bacteriol.* 187, 125–134; Fig. 1, modified.

(degradation of 2-hydroxybiphenyl). The N-terminal domains of these XylR-type activator proteins are responsible for the binding of their cognate aromatic effector, upon which the ATPase activity of the C-domain is exposed, resulting in transcription activation from the σ^{54}-dependent promoter.

W. Boos, H. Shuman **1998**, Maltose/maltodextrin system of *Escherichia coli*: transport, metabolism and regulation, *Microbiol. Mol. Biol. Rev.* 62, 204–229.

J.D. Gralla **1996**, Activation and repression of *Escherichia coli* promoters, *Curr. Opin. Genet. Dev.* 6, 526–530.

L. Reitzer, B.L. Schneider **2001**, Metabolic context and possible physiological themes of σ^{54}-dependent genes in *Escherichia coli*, *Microbiol. Mol. Biol. Rev.* 65, 422–444.

6.2.8.3 Two-component Signal Transduction Systems

The most adaptable eubacteria contain a large reservoir of genetic information encoding biochemical pathways, allowing them to respond instantaneously to rapidly changing environments. Eubacteria equipped with the genetic capability to respond to altered conditions are stimulated by specific signals. Recognition of specific signals and conversion of this information into specific transcriptional responses is the hallmark of signal transduction. One important mechanism found in all eubacteria is the *two-component signal transduction system*. In most cases, as the designation implies, it consists of two different proteins, one serving as the *sensor kinase* and the other as its cognate *response regulator*. Typically, the sensor kinases are integral membrane proteins and divisable into two functional domains, an N-terminal *sensor detection domain* followed by an *autokinase domain* consisting of a histidine-containing phosphotransferase subdomain and an ATP-binding subdomain (Fig. 6.25). Sensor detection domains are heterogeneous in size and

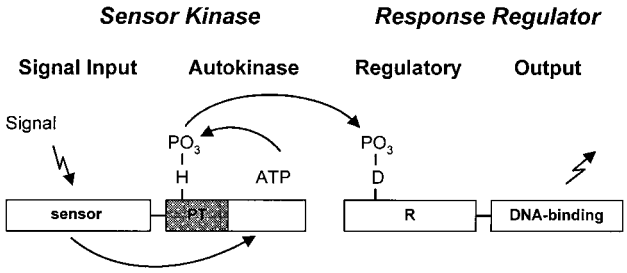


Fig. 6.25 Schematic diagram depicting the modular architecture of the two-component systems. Sensor kinases generally consist of a signal input domain (sensor domain) and an autokinase domain subdivided into a histidine phosphotransferase and an ATP-binding subdomain. Upon interaction of a stimulus with the input domain, autophosphorylation of an invariant histidine residue occurs. Response regulators also consist of two functional domains, a regulatory domain at the N-terminus followed by an output domain at the C-terminal half which is normally involved in DNA binding. Transfer of the phosphoryl group from the sensor kinase to the response regulator leads to the activation of the output domain enabling the protein to bind to its cognate sites on DNA. J.A.Hoch **2001**, *J. Bacteriol.* 183, 4941; Fig. 1.

amino acid sequence, reflecting the variety of stimuli detected. In contrast, the autokinase domains with which the response regulator must interact are of similar length and show many conserved amino acid motifs indicative of a common evolutionary origin. Thus, in integral membrane sensor kinases, which constitute the majority of sensor kinases, the cytoplasmic response regulator must find its partner among highly similar autokinase domains protruding internally from the cytoplasmic membrane. Response regulator transcription factors consist of two functional domains; and the N-terminal *regulatory domain* accepts phosphoryl groups at an invariant aspartate residue and regulates the activity of the C-terminal *DNA-binding domain*. The regulatory domain consists of about 120 amino acids and folds into a structure common to all domains of this type. The DNA-binding domains that determine the promoter specificity of the transcription factor are more heterogeneous in sequence and structure. While the vast majority of response regulators are transcription factors, it is important to point out that some response regulators do not have a C-terminal domain and others have an enzyme as the C-terminal domain. Examples of both may be found in chemotaxis signaling pathways. Two-component signaling systems do not exclusively occur in eubacteria, but have also been described in Archaea and eukaryotes, including *S. cerevisiae*, *Candida albicans*, *Neurospora crassa* and *Arabidopsis thaliana*. Bacteria such as *E. coli* and *B. subtilis* possess 30–40 different pairs of two-component systems, each dedicated to unique signals and genes. In addition to the classic two-component systems, a striking feature is the occurrence of a multistep His → Asp → His → Asp *phosphorelay* through three (or more) types of phosphotransfer signaling domains. These phosphorelay systems can be classified in two types: *intramolecular* and *intermolecular* phosphorelays. Intramolecular phosphorelays make use of one sensor kinase, where the phosphate is transferred within this polypeptide chain from histidine to aspartate and another histidine residue. The *E. coli* ArcAB phosphorelay will be discussed as an example. In intermolecular phosphorelays, the phosphate transfer from the sensor kinase to the response regulator involves several proteins. The *B. subtilis* sporulation phosphorelay will be presented as the best studied example.

Signal reception by the sensor kinase stimulates an ATP-dependent autophosphorylation at the conserved histidine residue. The phosphorylated sensor kinase then catalyzes a transphosphorylation of the cognate response regulator at its conserved aspartate residue, thereby rendering it functional, typically as a transcriptional regulator. Upon cessation of subsidence of signaling, both the sensor kinase and the response regulator undergo dephosphorylation, allowing signal decay and silencing of the system. Dephosphorylation is catalyzed either by the sensor kinase (acting as a bifunctional enzyme) or the autophosphatase activity of the response regulator or by a phosphatase depending on the system. Two classes of two-component systems have been distinguished: the 'classic' and the 'unorthodox' type, where the classic system of two-component systems predominates. This type of system is characterized by a one-step phosphotransfer reaction from the sensor to the response regulator protein. Unorthodox systems, also called phosphorelays, alternate between several histidine and aspartate residues (His-Asp-

His-Asp) and will be described in the next chapter. *E. coli* possesses some 30 two-component systems; and 34 have been postulated for *B. subtilis*. In the following, three classic two-component signal transduction systems will be described in detail, namely the EnvZ/OmpR system of *E. coli* involved in osmosensing, the VirA/VirG system of *A. tumefaciens* involved in sensing phenolic substances present in the exudat leaking from wounded plant cells and the very sophisticated chemotaxis system of *E. coli* allowing cells to move in gradients of attractants and repellents.

The EnvZ-OmpR Two-component System of *E. coli* is Involved in Osmoregulation
Osmoregulation in *E. coli* is essential for survival in rapidly changing environments such as fresh water with low solute concentration as well as in the gut of an organism with high solute concentration. When cells experience an increase in medium osmolarity (hyperosmotic shock), they first respond by rapidly enhancing intracellular potassium levels to restore negative turgor. Next, accumulation of other compatible solutes commences, including glycine betaine, taurine (an amino acid present in bile), choline (a precursor of glycine betaine) and trehalose, which do not interfere with intracellular metabolism, and potassium efflux systems are activated. In addition, the OmpF and OmpC porin (see Section 1.4) composition in the outer membrane is dramatically altered via activation of the EnvZ-OmpR two-component system. EnvZ is the inner membrane sensor kinase and *ompR* codes for the response regulator. The primary osmotic signal or natural ligand for EnvZ has not been identified. While it is possible that a distinct ligand exists, it is more likely that EnvZ is activated by a mechanical signal which may result in conformational changes or receptor dimerization. When the sensor domain interacts by the unknown signal, EnvZ is phosphorylated at histidine 243. It transfers this phosphoryl group to aspartic acid 55 of OmpR which then interacts with two different OmpR binding sites with a winged HTH DNA binding motif located within the C-terminal domain. These binding sites are located in the *ompF* and *ompC* promoter regions. The level of phosphorylation determines which porin gene is preferentially expressed. At low osmolarity, there are low levels of OmpR~P, resulting in activation of OmpF through binding to high affinity sites in the promoter region. High osmolarity results in higher levels of OmpR~P, leading to repression of *ompF* through binding to both low and high affinity sites, involving a proposed repression loop (see Fig. 6.26). In addition, phosphorylated OmpR binds to the three additional sites, C1 to C3, located upstream of *ompC*, leading to transcriptional activation. By binding to these sites, OmpR~P activates expression of a second gene, *micF*, located immediately upstream of *ompC* and transcribed in the opposite direction. The *micF* gene codes for a 92-bp transcript which exhibits partial complementarity to the 5′ end of the *ompF* transcript. When both RNAs anneal, translation of the *ompF* mRNA is prevented, suggesting that *micF* acts as anti-sense RNA. Phosphorylation and dephosphorylation of OmpR is controlled by EnvZ kinase and phosphatase activities, respectively, where the switch between these two enzymatic activities is modulated by osmolarity and some other environmental factors.

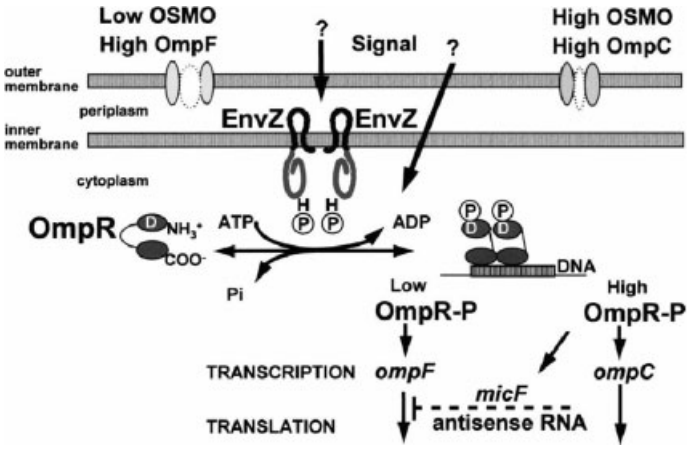


Fig. 6.26 Osmoregulation in *E. coli* via the EnvZ-OmpR two-component system. The total number of the outer membrane porins OmpF and OmpC is kept constant, but the relative number can vary. While OmpF forms a large pore that predominates at low osmolarity, OmpC forming a small pore takes over at high osmolarity. High osmolarity is sensed by EnvZ undergoing autophosphorylation and transfer of the phosphoryl group to an invariant aspartate residue in the N-terminal domain of OmpR. Phosphorylation leads to a conformational change allowing OmpR~P as a dimer to interact with binding sites located upstream of *ompF*, *ompC* and *micF*. At low osmolarity, there are low levels of OmpR~P which bind primarily to two of the four *ompF* upstream promoter binding sites resulting in transcriptional activation of *ompF*. At high osmolarity, more OmpR~P molecules are present which bind to all four binding sites resulting in a proposed loop and repression of *ompF*. On the contrary, binding to the three sites upstream of *ompC* leads to transcriptional activation of *ompC* itself and of the *micF* gene located upstream of *ompC* and transcribed in the opposite direction. The *micF* gene codes for a 92-bp RNA complementary to part of the *ompF* mRNA, acts as an anti-sense RNA and blocks its translation. L.A.Egger **1997**, *Genes Cell* 2, 167; Fig. 3.

Chemotaxis of *E. coli* is Regulated by a Complex Two-component System Like other motile bacteria, *E. coli* responds to chemical gradients by moving towards higher concentrations of attractants, lower concentrations of repellents and the optimal growth temperature, swimming at 10–20 body lengths per second. Cells rotate stiff flagellar filaments like a propeller and perform a biased random walk. This random walk consists of alternating runs (periods of forward movement) and tumbles (sudden reorientation) that arise from changes of flagellar rotation. When the flagella rotate counterclockwise (CCW), they form a bundle allowing the bacterium to swim in a more or less straight line at a roughly uniform speed. When one or more flagella rotate clockwise (CW), the bundle falls apart and the bacterium tumbles, randomly reorienting itself. The bacterial cell biases the random walk by modulating the duration of the run in response to measurements of chemoattractant concentration that are carried out at the cell surface. This behavior, termed *chemotaxis*, is mediated by a dedicated sensory system comprising five transmembrane chemoreceptors, histidine and aspartate kinases, a coupling protein (CheW) and two enzymes that mediate *sensory adaptation* by covalently

modifying the chemoreceptors. Homologs of these sensory components occur in virtually every motile eubacterium or archaeon investigated to date, making this type of sensory pathway one of the most prevelant in nature.

Five different integral membrane sensory proteins called MCPs (for *m*ethyl-accepting *c*hemotaxis *p*roteins) have been identified in *E. coli*, while *V. cholerae* inherits more than 40 MCPs. These are Tar (*t*axis for *a*spartate and *r*epellent), Tsr (*t*axis for *s*erine and *r*epellent), Trg (*t*axis for *r*ibose and *g*alactose), Tap (*ta*xis for oligo*p*eptides) and Aer for O_2 and cellular redox potential. These MCPs can sense a variety of compounds beyond those given, and in some cases, the chemoreceptors bind ligands directly (e.g., Tar binds aspartate, Tsr binds serine), whereas in other instances chemoreceptor interactions with small-molecule ligands are mediated by soluble periplasmic binding proteins (see Fig. 6.27). For example, Tar senses maltose as well as repellants such as the heavy metals, cobalt and nickel. Some bacterial species also contain cytoplasmic chemoreceptors which are active in clusters. The MCPs consist of α-helical coiled-coil structures and form a noncovalent complex with the sensor kinase CheA and the coupling protein CheW, which stabilizes the complex formed between the receptor and CheA (Fig. 6.27). Ligands bind to the four-helix bundle of the periplasmic domain at the interface between the two monomers of the dimer; and residues from both monomers are involved in the binding reaction. Ligand binding changes the interactions between the periplasmic domains of the MCPs and between the transmembrane four-helix bundle, resulting in a 1.4-Å piston-like movement of one transmembrane helix against another in the MCP dimer. The adaptor protein CheW links the MCPs to the cytoplasmic histidine kinase CheA and two response-regulators, CheY and CheB. CheY is a single-domain, flagellar motor-binding protein, while CheB has two domains, one of which functions as a methylesterase and controls adaptation of the MCPs. CheY~P interacts with the switch protein FliM on the flagellar motor and causes it to switch to CW rotation, resulting in cell tumbling and direction change. The CheY~P signal is terminated by dephosphorylation through CheZ. If CheY~P is unphosphorylated, it cannot interact with the flagellar motor which rotates CCW and the cell undergoes a run. The CheZ protein is a phosphatase which specifically dephosphorylates CheY, allowing the cell to return to the run.

There is a second level of regulation called *adaptation* which involves methylation and demethylation of the MCPs. The methyltransferase CheR continously adds methyl groups to the MCPs, using S-adenosylmethionine as a methyl donor. MCPs contain four to six glutamate residues subject to methylation. Demethylation of the MCPs occurs by the phosphorylated form of CheB, a second response regulator of the CheA kinase. The level of methylation of the MCPs (up to three methyl groups can be added per monomer) affects their conformation and controls adaptation to the stimulus. If the level of attractant remains high, the level of CheA~P remains low; and the cell swims smoothly and methylation of the MCPs increases until they are fully methylated. This in turn leads to an increase in the level of CheA~P and CheB~P and the cell tumbles. Demethylation of MCPs allows binding of the attractant and the cell resumes smooth swimming. If cells are ex-

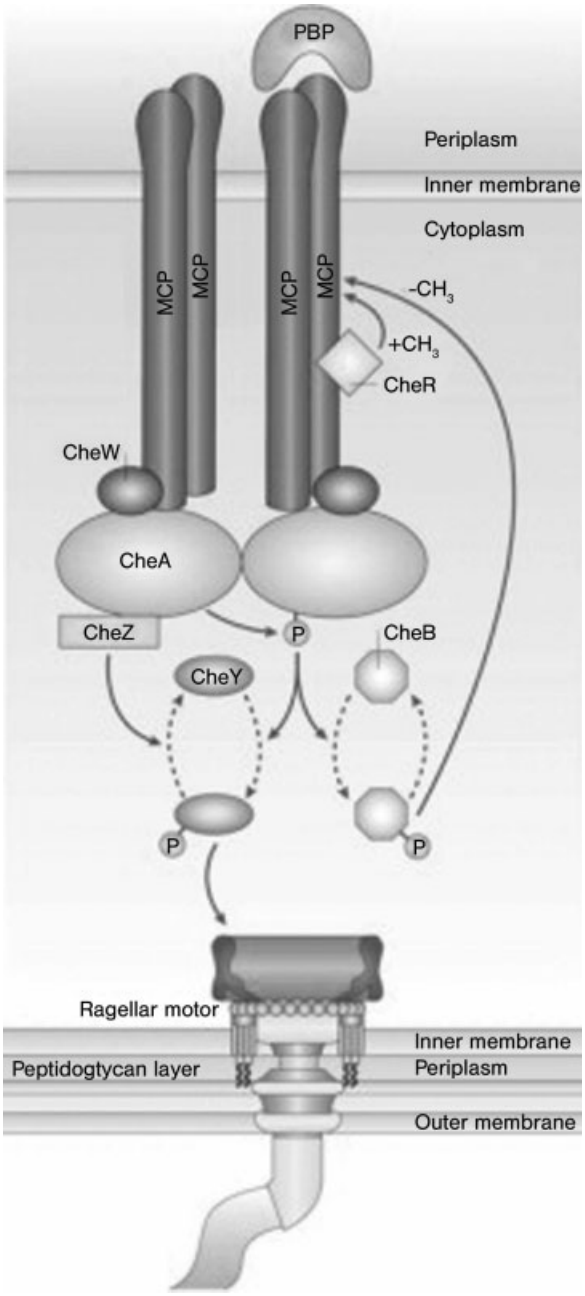


Fig. 6.27 Signalling pathways in bacterial chemotaxis. Dimeric MCPs form α-helical coiled-coil structures spanning the inner membrane. Chemical compounds (attractants and repellents) enter the periplasm through porins and interact either directly with the periplasmic domain of the appropriate chemoreceptor or indirectly through a *p*eriplasmatic *b*inding *p*rotein (PBP; MBP, RBP, GBP and DPP are *m*altose, *r*ibose, *g*alactose and *d*ipep*t*ide *b*inding *p*rotein, respectively). The CheW coupling molecule transduces the signals to the CheA sensor kinase which, after *trans*-phosphorylation, transfers the phosphate first to CheY and then to methylesterase CheB. CheY~P interacts with the flagellar motor to bring about a change in the direction. CheB~P competes with a constitutive methyltransferase, CheR, to control the degree of methylation of specific glutamates in the MCPs. This resets the signaling state of the chemoreceptors and allows them to adapt to the present concentration of attractant and to sense subsequent changes. Dephosphorylation of CheY~P is accelerated by the phosphatese CheZ. G.H. Wadhams **2004**, *Nat. Rev. Mol. Cell Biol.* 5, 1024–1037; Fig. 2. (This figure also appears with the color plates.)

posed to a repellent, the situation is opposite. Here, fully methlyated MCPs respond best to repellents.

Chemoreceptor complexes are localized predominantly at the cell poles. Studies of the chemotaxis machinery in *E. coli* have shown that chemoreceptors are organized into large membrane-associated, two-dimensional arrays. Each of the chemoreceptors assembles into a two-dimensional lattice containing hundreds or thousands of receptors, which are held together by bridging connections to CheA and CheW. The receptors form dimers and these dimers form trimers that can be of mixed receptor composition. These mixed receptor trimers bind CheA and CheW to form signaling teams. The mechanism that causes the chemoreceptors to congregate in clusters at the cell poles in not known. *E. coli* can detect and respond to extremely low concentrations of attractants, concentrations of less than 5 nM in the case of aspartate. They also sense gradients of attractants extending over five orders of magnitude in concentration (up to 1 mM aspartate).

The Arc Intramolecular Phosphorelay of *E. coli*

The Arc two-component signal transduction system is activated in response to anaerobic growth conditions and comprises the transmembrane sensor kinase, ArcB, and the response regulator, ArcA (Fig. 6.28). Upon a shift from aerobic to anaerobic conditions, ArcB undergoes autophosphorylation, presumably due to a change of redox signals in the membrane. The Asp-54 of the response regulator ArcA is phosphorylated either through the N-terminal histidine residue or after the phosphoryl group on phosphorylated ArcB has undergone intramolecular phosphorelay from His-292 to Asp-576 to His-717. The phosphoryl group of ArcA-P is unstable, with a half-life of 30 min. ArcA exists as a homodimer and ArcA~P multimerizes to form a tetramer of dimers which is composed of both ArcA~P and ArcA in a 1:1 ratio. The SixA phosphatase acts on the C-terminal histidine residue, but its *in vivo* function remains elusive. Phosphorylated ArcA positively and negatively regulates the transcription of a number of operons involved

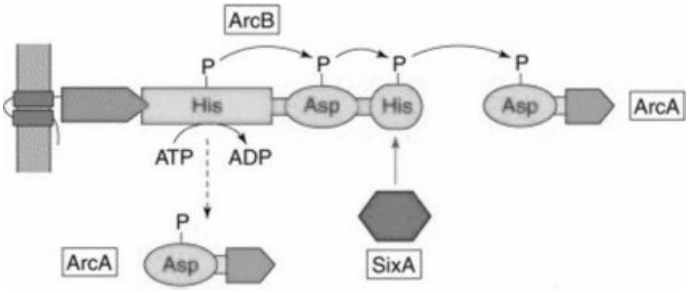


Fig. 6.28 The intramolecular ArcAB phosphorelay of *E. coli*. Phosphorylation of the ArcA response regulator can occur either directly by the N-terminal histidine or through in internal phosphorelay. SixA is a phosphohistidine phosphatase which can dephosphorylate the C-terminal histidine residue. A.-L. Perraud **1999**, *Trends Microbiol.* 7, 115; Fig. 2a.

either directly or indirectly in cellular metabolism under both aerobic and anaerobic conditions. Recently, it was reported that ArcA-P binds to *oriC*, thereby specifically inhibiting *in vitro E. coli* chromosomal replication.

The Intermolecular Phosphorelay of Sporulation Control of *B. subtilis*

Under conditions of nutrient depletion at high cell densities, the response of last resort for *B. subtilis* is to initiate the process of endospore formation. Initiation of sporulation involves an intermolecular phosphorelay system, leading to the phosphorylation of the response regulator Spo0A. Multiple kinases (five have been identified so far, but only two of them, KinA and KinB, have been studied in detail) provide signal input into the system through an autophosphorylation reaction, with subsequent transfer of a phosphoryl group to the Spo0A transcription factor via the Spo0F response regulator and the Spo0B phosphotransferase intermediates (Fig. 6.29). The use of a multicomponent system, in place of the classic two-component system, was proposed to provide multiple entry levels to negative regulators for controling the flow of phosphoryl groups in the system and the ultimate production of Spo0A~P. Negative regulation is carried out through controlled dephosphorylation at the level of Spo0F~P and Spo0A~P response regulators. The phosphorylation level of Spo0A is specifically and directly modulated by the Spo0E phosphatase in response to signals that remain unknown. Spo0F~P is the target for the RapA and RapB phosphatases. These response regulator aspartyl phosphate phosphatases provide access for negative signals to influence the cell's

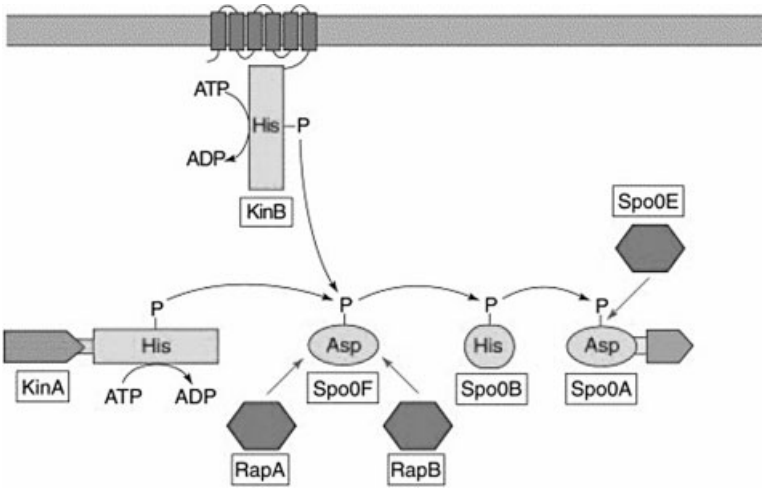


Fig. 6.29 The intermolecular phosphorelay of *B. subtilis* controls initiation of sporulation. Upon receiving a signal, the membrane-bound sensor kinase KinB or the cytoplasmic kinase KinA undergo autophosphorylation and phosphate is transferred first to Spo0F, then from Spo0F to the phosphotransferase Spo0B, and finally from Spo0B to Spo0A. The two phosphatases RapA and RapB negatively regulate Spo0F, while the phosphatase Spo0E can dephosphorylate Spo0A~P. A.-L. Perraud **1999**, *Trends Microbiol.* 7, 115; Fig. 1b.

decision of whether to initiate the sporulation process or to continue with vegetative growth.

6.2.8.4 Quorum Sensing

Bacteria actively ''speak'' to one another by sending and receiving messages in the form of chemical signals. In a process designated as *quorum sensing*, bacteria measure the concentration of these signaling molecules, termed *autoinducers* or *peptide pheromones*. Here, the concentration of the autoinducer reflects the number of bacterial cells and the perception of a threshold level of the signaling molecule indicates that the population has reached the critical quorum. This in turn initiates an altered pattern of gene expression. Quorum sensing constitutes a mechanism for multicellular behavior in prokaryotes and is now known to regulate a variety of processes, including production of secondary metabolites in *Erwinia* and *Serratia*, root nodule symbiosis in *Rhizobium* species, bioluminescence in *V. fischeri* and *V. harveyi*, virulence factor expression and biofilm development in *P. aeruginosa* associated with a wide variety of chronic infections, competence in *B. subtilis* and *S. pneumoniae* and conjugation, as well as individual survival strategies such as induction of stationary phase responses and motility. Cell-to-cell signaling does not solely occur at high cell densities, and the term ''quorum sensing'' is now applied to any bacterial intercellular communication that depends on small diffusible signaling molecules. While cell-to-cell signaling normally occurs between cells of the same species (intraspecies), the signals can also be detected by another species (interspecies). Two major chemical classes of autoinducers have been identified: (a) *fatty acid derivatives* (*N-a*cyl-*h*omoserine *l*actones; AHLs) used by Gram-negative bacteria and (b) short *oligopeptides*, sometimes modified, commonly utilized by Gram-positive bacteria. An exception is the universal pheromone LuxS which has been found in both groups of bacteria. The autoinducers are secreted outside the cells, followed by uptake by the bacteria where they trigger gene expression. While the fatty acid derivatives diffuse through the two membranes into the cytoplasm, the short peptides are taken up by two-component signal transduction systems (see above).

6.2.8.4.1 Quorum Sensing in Gram-negative Bacteria

In more than 30 Gram-negative bacteria, quorum sensing is mediated by lipid signaling molecules that are chemical derivatives of AHL autoinducers. AHLs are synthesized by AHL synthases (e.g., LuxI) and are sensed by the response regulator family of transcription factors known as R proteins (e.g., LuxR; Fig. 6.30). Intracellular accumulation of a sufficient oncentration of the cell-permeable AHL leads to the activation or repression of specific target genes. This is accomplished by the binding of AHL to the cognate R protein.

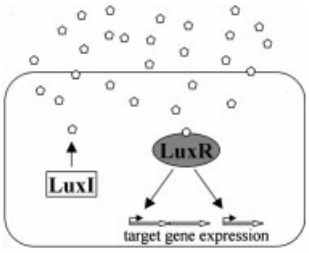


Fig. 6.30 The quorum sensing system of Gram-negative bacteria. The LuxI-type protein catalyzes synthesis of an AHL autoinducer which is secreted into the medium. The more cells are present in a given volume, the more autoinducer is synthesized and secreted. This increases the chance that autoinducer is taken up by cells where is binds to the LuxR transcriptional activator, which stimulates transcription of the target genes. M.E. Taga, B.L. Bassler **2003**, *Proc. Natl. Acad. Sci. USA* 100, 14549–14554; Fig. 1.

The Lux System Encoded by *V. fischeri*

V. fischeri is a marine bacterium that is nonluminescent when free-living in seawater. When these cells are grown to high cell densities in the laboratory, the culture bioluminesces with a blue-green light. *V. fischeri* is able to form symbiotic relationships with some fish and squid species. The best studied example of such a symbiosis is that with the small squid *Euprymna scolopes*. This squid lives in shallow water, is a nocturnal feeder and has a bioluminesent appearance in dark environments due to the maintenance of high-densitiy *V. fischeri* populations (10^{10}–10^{11} cells ml^{-1}) in a specialized light organ. *E. scolopes* uses the bioluminescence for counter-illumination to eliminate the visible shadow created by moonlight and starlight that it projects downwards. In return, *E. scolopes* provides the bacterial cells with nutrients. When the host buries itself in the sand for daytime quiescence, it expels ~95% of its light-organ symbionts into the surrounding environment. During the day, the bacteria that remain in the light organ multiply to repopulate it so that the animal has a full complement of symbionts when it emerges from the sand. The symbiosis between the squid and the bacterium is an interesting example of an alliance between an animal and a microorganism.

The bioluminescence gene cluster of *V. fischeri* consists of eight *lux* genes (*luxA–E, luxG, luxI, luxR*) arranged in two bi-directionally transcribed operons. While the *luxI* gene codes for the AHL autoinducer, *luxR* specifies the R protein. The genes *luxA* and *luxB* code for the two subunits of the heterodimeric luciferase enzyme. This protein catalyzes the oxidation of aldehyde and reduction of flavin mononucleotide resulting in a long-chain fatty acid, water, a flavin mononucleotide and the simultaneous liberation of excess free energy, evident as blue-green light. The genes *luxC–luxE* code for proteins involved in the synthesis of the aldehyde substrate and *luxG* is probably a flavin reductase. At low cell density, transcription of the *lux* genes is weak and no light is produced. When the cell density increases and a threshold concentration of AHL [in this case, *N*-(3-oxohexanosyl)-L-homoserine lactose; OHHL] is reached, bioluminescence commences. *V. fischeri* cells express the *luxI* gene at a basal level at low cell densities, resulting in a low concentration of OHHL in the medium.

Quorum Sensing in *P. aeruginosa* *P. aeruginosa* is an important opportunistic human pathogen often associated with infections of immunocompromised patients and in patients suffering with AIDS, cystic fibrosis or severe burns. The basis of pathogenicity is related to its ability to produce and secrete a variety of extracellu-

lar virulence factors, such as elastase (one of the most abundantly secreted proteases, with high activity and broad substrate specificity), exotoxin A, pyocyanin (belongs to the group of siderophores supposed to be involved in a reduction mechanism capable of recovering iron from transferrin) and hemolysin, all capable of causing extensive tissue damage. Regulation of the genes coding for these exoproteins occurs by two different quorum sensing systems. The first system comprises LasI and LasR. LasI is responsible for the synthesis of *N*-(3-oxododecanoyl)-L-homoserine lactose (OdDHL), whereas LasR acts as a transcriptional activator. An increase in LasI production leads to a rapid increase in OdDHL production which activates LasR, resulting in a further increase in the amount of LasI, resulting in an autoregulatory loop.

The second quorum sensing system is controlled by RhlI (for *rh*amnol*i*pid) and RhlR. RhlI directs the synthesis of *N*-butanoyl-L-homoserine lactone (BHL), which interacts with the transcriptional activator RhlR. One of the RhlR-BHL-activated target genes is the *rhlAB* operon which codes for a rhamnosyl transferase. This enzyme is involved in the production of rhamnolipid biosurfactants which reduce the surface tension and allows *P. aeruginosa* to swarm over semi-solid surfaces. In addition, the RhlRI quorum sensing system is required for enhanced expression of other proteins, including pyocyanin, hydrogen cyanide, alkaline protease and elastase. As already described for the LasRI system, transcription of *rhlI* is increased by RhlR-BHL, creating a further autoregulatory loop. In addition, the RhlRI system is functionally dependent upon the LasRI system, since transcriptional activation of *rhlR* is dependent upon LasR-OdDHL.

Conjugal Transfer of the Ti Plasmid is Regulated by Quorum Sensing in *A. tumefaciens* *A. tumefaciens* is a pathogen that induces crown gall tumors in plants by transferring oncogenic DNA from its conjugative Ti plasmid to the plant genome (see Section 10.2.1). The transferred DNA directs the synthesis of opines which are used by *A. tumefaciens* cells, growing in the tumor, as a nutrient source and as a chemical signal to induce conjugal transfer of the Ti plasmid to bacterial cells present within the tumor and not containing this plasmid. Opines are modified amino acids and Ti plasmids are classified upon the opine genes they carry. Examples are octopine, nopaline, agropine and agrocinopine. When the tumor cells produce agrocinopine, this compound is taken up by the bacterial cells, where it interacts with the transcriptional repressor AccR and causes dissociation from its operator. This in turn causes expression of the transcriptional activator gene *traR*. To become active as an activator, TraR has to bind the autoinducer *N*-(3-oxooctanoyl)-L-homoserine lactose (OOHL) which is synthesized by TraI. TraR-OOHL activates transcription of the *tra* and *trb* genes required for conjugal transfer. Another level of regulation is exerted by the 11-kDa anti-activator TraM protein. This small protein binds to the C-terminal domain of TraR in an OOHL-independent manner. The physiological purpose of the TraM-TraR interaction may be to repress TraR-mediated activation of conjugation at low cell densities and/or to reduce the quorum sensing system. Both interpretations are corroborated by the finding that expression of *traM* is positively regulated by TraR, creating another

regulatory loop. Why is conjugal transfer of the Ti plasmids regulated by use of a quorum sensing system? There are data suggesting that transfer to recipient cells is rather inefficient and the high number of donor cells ensures successful transfer.

S. Swift, et al. **2002**, Quorum sensing as a population-density-dependent determinant of bacterial physiology, *Adv. Microbiol. Physiol.* 45, 199–269.
N.A. Whitehead, et al. **2001**, Quorum-sensing in Gram-negative bacteria, *FEMS Microbiol. Lett.* 25, 365–404.
H. Withers, et al. **2001**, Quorum sensing as an integral component of gene regulatory networks in Gram-negative bacteria, *Curr. Opin. Microbiol.* 4, 186–193.

6.2.8.4.2 Quorum Sensing in Gram-positive Bacteria

In Gram-positive bacteria, quorum sensing signaling molecules are generally small peptides termed ''peptide pheromones'' or ''autoinducing peptide'' (AIP) as cell-to-cell comunication signals to mediate quorum sensing (Fig. 6.31). These small peptides are either unmodified or more often are post-translationally modified. In the case of the *S. aureus*-encoded AgrD pheromone, this modification was shown to be an intramolecular thiolactone bond linking the thiol group of a central cysteine to the C-terminal carboxyl group. In contrast, the ComX pheromone encoded by *B. subtilis* has an extra moiety added post-translationally and was proposed to be a farnesylation. These peptides are secreted either via a dedicated

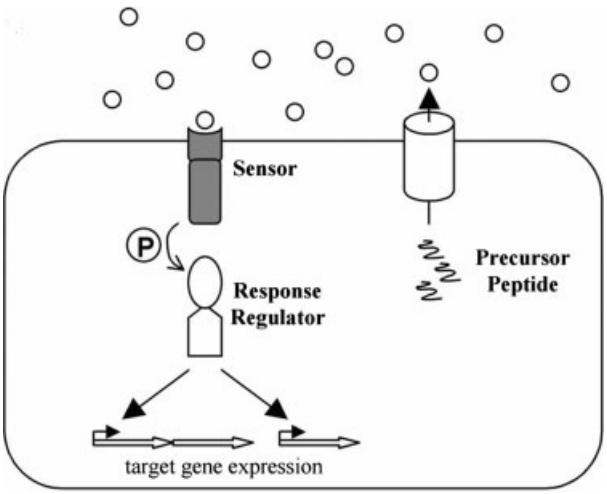


Fig. 6.31 The quorum sensing system of Gram-positive bacteria. A precursor peptide called autoinducing peptide (AIP) or ''peptide pheromone'' is synthesized within the cytoplasm, cleaved, modified and exported. AIPs specifically interact with the external domains of a membrane-bound two-component sensor kinase, which stimulates its kinase activity. The phosphate is transferred to the response regulator which binds to DNA and alters gene expression. M.E. Taga, B.L. Bassler **2003**, *Proc. Natl. Acad. Sci. USA* 100, 14549–14554; Fig. 1.

ATP-binding cassette (ABC) transporter (see Section 1.2) or, when equipped with a signal peptide, via the general Sec pathway (see Section 8.4.1). These peptide pheromones are either taken up by an oligopeptide permease or are sensed by a two-component signal transduction system (see Section 6.2.8.3). Several Gram-positive quorum sensing systems have been extensively studied. I will describe the systems controling competence and sporulation in *B. subtilis* and virulence in *S. aureus*.

ComX and CSF Mediate the Quorum Response in *B. subtilis* In *B. subtilis*, two signaling peptides, ComX and CSF (for *c*ompetence and *s*porulation *f*actor), regulate the activity of the transcription factor ComA (Fig. 6.32). The 55-amino-acid ComX pheromone is secreted to the outside of the cells via the Sec pathway (see Section 8.4.1) and as the cell density increases, ComX reaches a concentration that binds and activates the sensor kinase, ComP. ComP~P then transfers the phosphate to ComA, generating ComA~P, which in turn activates transcription of those genes involved in the quorum response (Fig. 6.32). The main function of the ComX pheromone appears to monitor cell density. The second regulator of ComA activity is CSF, encoded by the *phrC* gene. The 40-amino-acid CSF is also secreted via the Sec pathway (see Section 8.4.1). When the active pentapeptid (ERGMT) CSF reaches a critical concentration, it is transported back into the cells by an oligopeptide permease (the Opp system). Inside the cells, it appears to bind to different in-

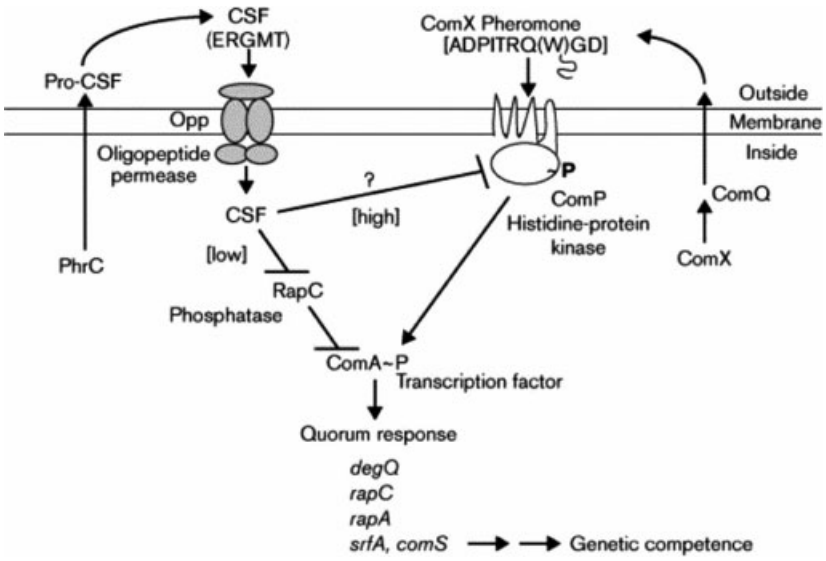


Fig. 6.32 Two extracellular signaling peptides mediate quorum sensing in *B. subtilis*. The *phrC* gene codes for the CSF (*c*ompetence and *s*porulation *f*actor) which inhibits the RapC phosphatase. The ComX pheromone binds to the ComP sensor kinase, which, after autophosphorylation, transfers the phosphate to the ComA transcription factor. See text for details. B.A Lazazzera **2000**, *Curr. Opin. Microbiol.* 3, 177–182; Fig. 1.

tracellular receptors, depending on the concentration. At low concentration (1–5 nM), CSF stimulates the activity of ComA apparently by inhibiting the activity of the aspartyl phosphate phosphatase RapC. At higher concentrations (>20 nM), CSF interacts with an as yet unidentified receptor, possibly the sensor kinase ComP, to inhibit the transcription of the ComA-controlled genes. In addition, CSF, at high concentrations, inhibits another aspartyl phosphatase, RapB, and thereby stimulates sporulation. Why, in *B. subtilis*, is cell density monitored by two different signaling molecules? It has been suggested that, while the ComX pheromone measures cell density, CSF signals starvation when expression of its gene *phrC* is enhanced as cells enter the stationary phase.

AgrD of *S. aureus* The *agr* (for *a*ccessory *g*ene *r*egulator) of *S. aureus* controls the expression of several extracellular virulence factors, including α- and β-toxins, DNase and toxic shock syndrome toxin-1. This global regulator is activated by a secreted autoinducing peptide. The *agr* locus consists of the four genes *agrABCD*, where *agrA* codes for the cytosolic regulator, *agrB* for the transmembrane and *agrC* for a sensor kinase of the quorum sensing system. The product of the *agrD* gene is a 46-amino-acid propeptide that is processed to the final octapeptide and secreted through AgrB. The mature autoinducing peptide contains a thiolactone ring that is absolutely required for full signaling activity. It is thought to interact with the transmembrane of AgrC resulting in autophosphorylation on the invariant histidine residue. *Trans*-phosphorylation of the AgrA transcriptional regulator of this two-component system leads to activation of RNAIII, which in turn functions as an effector to activate the expression of several secreted virulence factors. *S. aureus* strains can be placed in at least four different groups secreting slightly different autoinducers. But all of them carry a cysteine residue at position five and this residue is involved in the formation of an intramolecular thiol ester linkage between the sulfhydryl group and the C-terminus.

Some of the peptides have additional biological activities, such as the ability to lyse target cells. The best studied peptide of this group is the cytolysin produced by the human pathogen *E. faecalis*. This cytolysin is composed of two subunits, $CylL_L''$ (38 amino acids) and $CylL_S''$ (21 amino acids), which together form a complex that creates pores in the membranes of target eukaryotic cells. In the absence of target cells, basal levels of $CylL_L''$ and $CylL_S''$ are secreted and form inactive complexes. If target cells are present, $CylL_L''$ preferentially binds to the target cell membranes. Free $CylL_S''$ is taken up by the bacterial cells where it is assumed to bind to the repressor of the *cyl* operon (a complex of CylR1 and CylR2) to stimulate transcription. This in turn increases the amount of $CylL_L''$ and $CylL_S''$ secreted and these form complexes creating pores in the eukaryotic cell membranes. Therefore, $CylL_L''$ senses host cells and then targets their destruction.

C.M. Waters, B.I. Bassler **2005**, Quorum sensing: cell-to-cell communication in bacteria. *Annu. Rev. Cell Dev. Biol.* 21, 319–346

6.3 Post-transcriptional Regulation

Post-transcriptional regulation can occur by one of the following three different mechanisms.

Protein-mediated Regulation

Three different regulators can be distinguished base on the number of target sites: regulators with a single control target (e.g., autoregulation by ribosomal proteins in *E. coli*), regulators with several control targets (e.g., TRAP in *B. subtilis*) and so-called global regulators with many targets (e.g., CsrA).

Regulatory RNA-mediated Control, Also Called Riboregulation

E. coli codes for more than 60 small ncRNAs (for *n*on*c*oding RNAs) which have been shown to be involved in:

- secretion: 4.5S RNA (see Section 8.4.3)
- cleavage of RNA: RNase P (see Section 6.3.3)
- rescue of stalled ribosomes: tmRNA (see Section 6.4.2)
- gene regulation: regulatory RNAs interact with:
  - *cis*-encoded: target and regulatory RNA at the same location, e.g., *hok-sok* system
  - *trans*-encoded, e.g., *E. coli* MicC and MicF.

Riboswitches

Riboswitches are secondary structures within the 5′ untranslated region of some transcripts to which a metabolite will bind, thereby turning off (in most cases) or on (rarely) expression of the downstream structural genes. They act by:

- transcription attenuation
- controling translation initiation
- controling mRNA processing.

6.3.1 Protein-mediated Regulation

Autoregulation of Ribosomal Proteins

The synthesis of many ribosomal proteins is regulated to prevent the accumulation of free ribosomal proteins in the cytoplasm. In *E. coli*, genes coding for these proteins generally form operons. Expression of these operons is often under autogenous negative translational regulation where one of the operon's products binds to its own mRNA to repress further translation. In general, these translational repressors interact with the leader region of their own mRNA to directly inhibit translation of the first cistron and repression is transmitted to the downstream

genes by translational coupling (see Section 6.4.2). Most of these translational repressor proteins are also able to bind rRNA directly. These autoregulatory proteins preferentially bind to their primary target, rRNA, and under conditions where rRNA targets are not available for binding, they interact with the operator sites on their own mRNA to block translation. It has been suggested that these proteins bind to structurally similar mRNA and rRNA sites, termed *mimicry*. This mimicry hypothesis explains how competition between rRNAs and mRNAs for binding ribosomal proteins guarantees the coordinate expression of the ribosomal protein genes in response to cellular rRNA levels. Examples for dual function ribosomal proteins are S4, S7 and S8 for the 30S ribosomal subunit and L1, L10 and L20 for the 50S subunit. In all these cases, the regulatory protein binds to just one single operator site.

TRAP in *B. subtilis*

TRAP is a protein (encoded by *mtrB*) which forms an oligomeric structure mainly involved in transcription attenuation of the *trp* operon (see Section 6.2.7). In addition, this protein regulates expression of three additional operons at the level of translation (Fig. 6.33). Two target genes are *trpG* and *trpE*. In the case of *trpE*, TRAP binding promotes the formation of an RNA structure that sequesters the SD sequence. In the case of the *trpG* transcript, TRAP interacts to a sequence that carries the SD sequence. In addition to the tryptophan biosynthetic genes, TRAP regulates expression of a potential tryptophan transport gene (*yhaG*) and probably that of *ycbK*. In the case of *yhaG*, TRAP blocks ribosome access. In summary, the

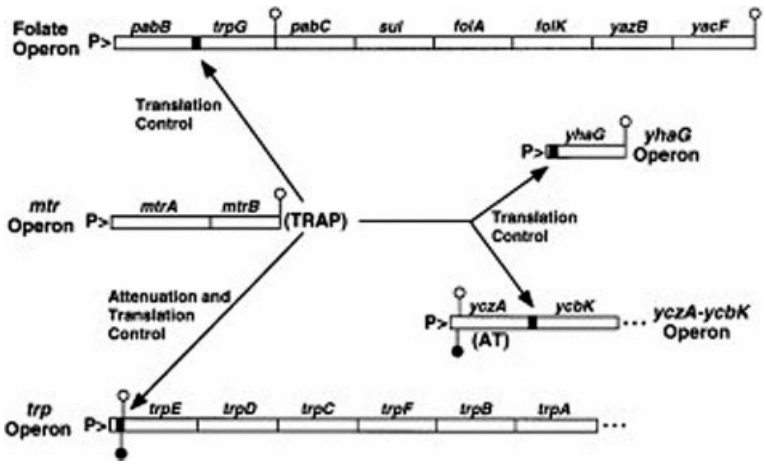


Fig. 6.33 The TRAP protein controls expression of four different operons. The promoters of the different operons are indicated by a P followed by an arrow, the TRAP binding sites by a black box. The location of terminators (open circle) and anti-terminators (filled circle) is indicated. P. Babitzke, P. Golnick **2001**, *J. Bacteriol.* 183, 5795–5802; Fig. 1.

regulatory TRAP protein negatively regulates the expression of genes located in five different operons.

The Global Regulator CsrA

The *c*arbon *s*torage *r*egulator *A* (CsrA) is a central component of the global regulatory system Csr responsible for repression of a variety of stationary-phase genes. CsrA, a 61-amino-acid homomeric protein, negatively regulates gluconeogenesis, glycogen biosynthesis, catabolism and biofilm formation. In addition, CsrA can activate glycolysis, acetate metabolism and flagellum biosynthesis. CsrA acts post-transcriptionally by repressing gene expression of some essential enzymes. It does so by destabilizing target mRNAs by binding in a region between the –18 and +31 nucleotides of the coding region, including the ribosome-binding site. This leads to a downregulation of translation of these genes and increases the instability of these target mRNAs, which may decay by various mechanisms. These include a direct attack at the CsrA-mRNA complex or an alteration of the transcript conformation followed by ribonuclease attack. Intracellular levels of CsrA are regulated by two small RNAs, CsrB and CsrA, acting as antagonists by sequestering CsrA and preventing its binding to target mRNAs (see below).

6.3.2 Riboregulation

While the classic RNA molecules mRNA, tRNA and rRNA function in the translation process, another group of RNAs, called small ncRNAs, carry out diverse functions in *E. coli* and other eubacteria. The function of these small ncRNAs is diverse and *E. coli* small ncRNAs act: (a) to change the conformation of target mRNAs (DsrA, RprA), (b) block mRNA by occlusion of Shine–Dalgarno sequences (MicF, MicC, OxyR, RyhB), (c) degrade target mRNAs (DsrA, Oop, RyhB) and (d) titrate specific protein factors (OxyS, CsrB), sometimes in combination. A common theme is that many are involved in the regulation of specific target genes at the post-transcriptional level via complementary base pairing to their cognate mRNAs. A few exert different functions, such as the tmRNA (involved in the release of stalled ribosomes and tagging polypeptides for degradation; see Section 6.4.2), the 4.5S RNA (part of the signal recognition particle involved in targeting integral membrane proteins to the inner membrane; see Section 8.4.3) and RNase P RNA (RNA component of RNase P involved in processing pre-tRNA and some other RNAs). But most of the more than 60 small ncRNAs described in *E. coli* so far function as genetic regulators, a process termed *riboregulation*. These small regulatory RNAs act in one of two general ways: pairing with another polynucleotide strand, usually a mRNA, or acting as a molecular decoy for proteins. By pairing to a target RNA molecule, the regulatory RNA can change its structure and/or stability, resulting in inhibition or promotion of ribosome binding and ultimately changing the translational efficiency. If the small ncRNA acts as a molecular decoy, its structure and/or sequence resembles that of the proteins's alternative tar-

get and the regulatory RNA may compete for the protein binding to that target. These regulatory RNAs seem to fine-tune cellular responses to stress conditions, integrating environmental signals into global regulation. Some of them affect more than one target gene. Up to the beginning of 2004, a total of 62 sRNAs has been described. Several examples will be discussed.

DsrA Changes the Conformation of Target RNAs

DsrA RNA is a small (87 nucleotides) multifunctional genetic regulator of *E. coli*. DsrA RNA modulates the levels of the two global transcriptional regulators, RpoS and H-NS, and acts by sequence-specific RNA–RNA interactions to enhance the translation of *rpoS* RNA and to stabilize the *rpoS* transcript. In the case of H-NS, DsrA binds the *hns* transcript by specific base-pairing interactions and blocks *hns* translation. The first stem-loop region of DsrA melts out to contact *rpoS* mRNA, whereas the second stem-loop region basepairs with *hns* mRNA. This conformational change within DsrA acts to switch the translation state of two different target transcripts. In addition, DsrA acts as an acid resistance regulator in *E. coli* (see Section 9.8.1).

MicF and MicC Block mRNA by Occlusion of Shine–Dalgarno Sequences

Some ncRNAs act as antisense riboregulators and exert their function by forming base pairs with mRNAs that are generally encoded at separate loci. The net effect can be the up- or downregulation of target genes. Examples of small ncRNAs that block ribosome accessibility to target mRNAs and thereby decrease gene expression are MicF and MicC. The MicF RNA is encoded divergently from the gene encoding the OmpC porin and represses expression of OmpF, another porin, and its expression is increased at low temperatures and in minimal medium (Fig. 6.34A).

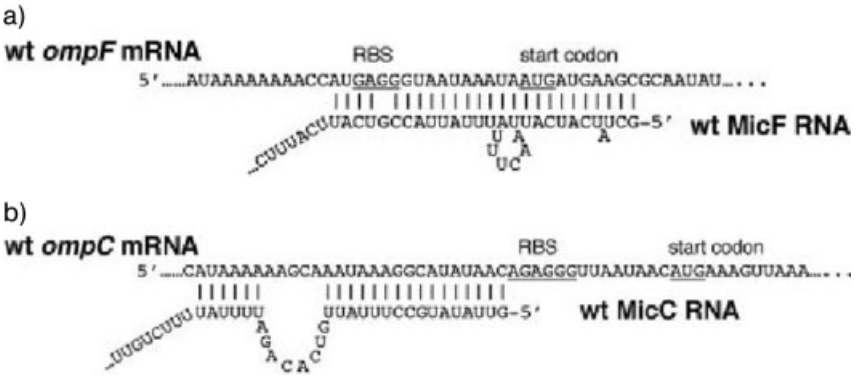


Fig. 6.34 Proposed formation of *micF-ompF* (A) and *micC-ompC* (B) duplexes. Base-pairing between the leader region of *ompF* and *micF* and *ompC* and *micC* RNA prevents binding to the ribosomal 30S subunit to the RBS. S. Chen, et al. **2004**, *J. Bacteriol.* 186, 6689; Fig. 3.

Twenty-two nucleotides at its 5′ end have the potential to form base pairs with the leader sequence of the mRNA coding for OmpC. MicC is a ~100-nucleotide RNA, this RNA inhibits ribosome binding to the *ompC* transcript leader and requires the Hfq RNA chaperone for its function (Fig. 6.34B).

OxyS RNA Inhibits Translation of Different mRNAs

OxyS RNA is induced in response to hydrogen peroxide stress and inhibits the translation of two target genes, *fhlA* (codes for the transcriptional activator of the formate regulon) and *rpoS*. The 109-nucleotide OxyS RNA basepairs near the Shine–Dalgarno sequence on the *fhlA* mRNA and blocks ribosome binding (Fig. 6.35). This RNA–RNA interaction is aided by the Hfq protein. For optimal OxyS RNA inhibition of *fhlA* translation, an additional basepairing is required with a region within the coding sequence. The second well studied target of OxyS RNA is the *rpoS* mRNA, where *rpoS* codes for an important alternative sigma factor involved in stationary phase and during general stress response. Here, too, there are regions of potential basepairing between OxyS RNA and *rpoS* mRNA, suggesting translational inhibition.

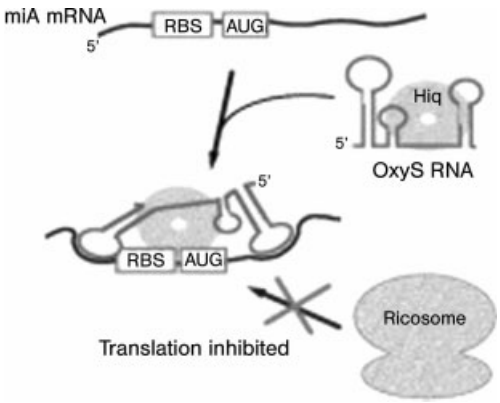


Fig. 6.35 OxyS RNA inhibits translation of the *fhlA* mRNA. The OxyS RNA interact with the Shine–Dalgarno sequence of the *fhlA* mRNA preventing its translation. This reaction is favored by the RNA-binding protein Hfq. K.M. Wassarman **2002**, *Cell* 109, 141; Fig. 1.

Antisense RNA Regulating Toxin Synthesis in the Plasmid R1

In low-copy-number plasmids, stable inheritance depends on a genetic system that actively prevents the appearance of plasmid-free progeny by killing plasmid-free progeny cells, also called post-segregational killing. In most cases, this genetic system codes for two proteins, a toxin and an antitoxin (also called antidote). The antitoxins neutralize their cognate toxins by forming tight complexes with them. Since the antitoxins are labile and the toxins stabile, continued synthesis is required for survival of the cells. If the plasmid is lost from a cell, the antitoxin fails to be synthesized, leading to the accumulation of free toxin that will kill the cell. Another type of toxin–antitoxin system uses an antisense RNA that inhibits translation of toxin-encoding mRNAs. While the antisense RNAs are unstable,

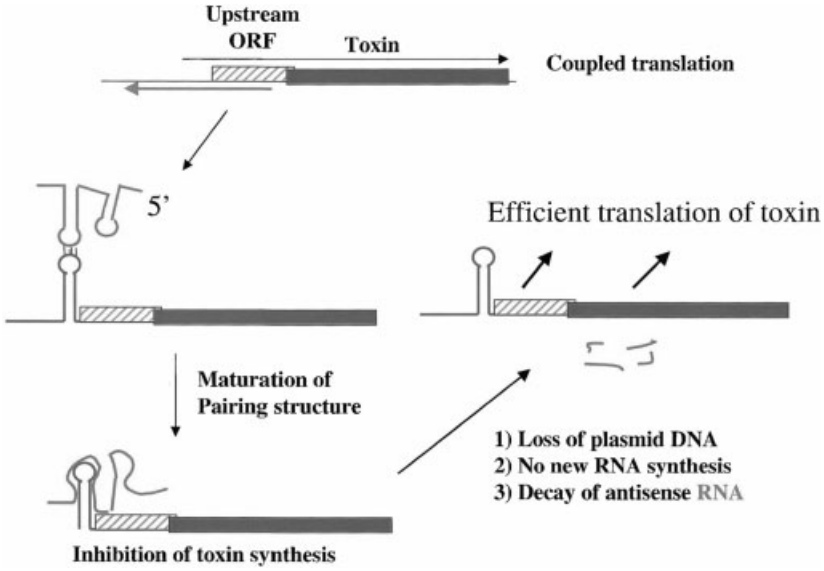


Fig. 6.36 Antisense RNA prevents translation of a toxin-encoding mRNA. Translation of the toxin mRNA is prevented by an antisense RNA (both coded for by a low-copy-number plasmid) exhibiting partial complementarity to the leader region of the mRNA. Both RNAs first interact via loops forming the kissing complex followed by maturation into a more extended double-stranded RNA structure. Loss of the plasmid results in the appearance of free toxin mRNA due to the decay of the antisense RNA which will be translated, and the toxin will kill the cell. S. Gottesman **2002**, *Genes Dev.* 16, 2829; Fig. 2.

the mRNAs coding for the toxins are rather stable. One of these systems is *hok*/*sok*, encoded by the plasmid R1. The genetic organization of the *hok*/*sok* system is shown in Fig. 6.36. The *hok* gene (for *ho*st *k*illing) codes for the membrane-associated toxin of 52 amino acid residues which causes irreversible damage to the cell membrane. The *sok* gene (for *s*uppressor of *k*illing) codes for the 64-nucleotide antisense RNA complementary to the *hok* mRNA region. As already mentioned, the toxin-encoding mRNA is rather stable, while the *sok* antisense RNA is unstable. Interaction between the two RNAs is initiated by loop pairing designated as a kissing complex which matures into a more extended interaction (see Fig. 6.36) preventing translation of the *hok* mRNA.

Regulation of CsrA Activity by CsrB Regulatory RNA

As mentioned above, CsrA acts as a global regulator in *E. coli* and exhibits a broad phylogenetic distribution in the eubacteria, repressing stationary-phase genes in *P. fluorescens* and genes involved in plant pathogenesis in *E. carotovora* and regulating genes involved in mucosal invasion by *S. enterica*. Its level is regulated by two small ncRNAs, CsrB and CsrC. CsrA binding to both CsrB and CsrC is mediated by a highly repetitive sequence element, 5′-CAGGA(U,C,A)G-3′, located

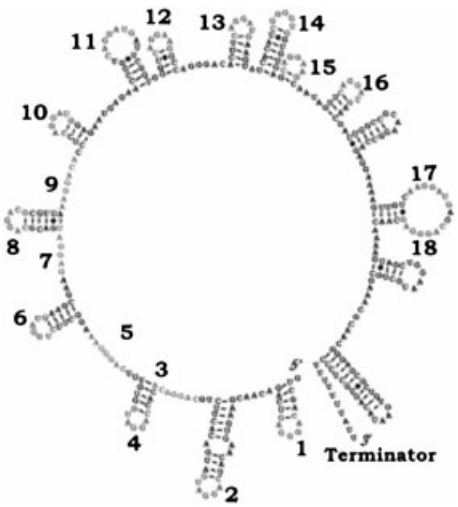


Fig. 6.37 Proposed structure of the CsrB RNA. This structure contains 18 repeated structures numbered 1 to 18 which may facilitate binding of CsrA. T. Romeo **1998**, *Mol. Microbiol.* 29, 1321–1330; Fig. 1. (This figure also appears with the color plates.)

in the loops of predicted CsrB/C hairpins (Fig. 6.37). CsrA activity in the cell is controlled through an equilibrium that exists between the free and the CsrB/C-bound protein. According to this model, accumulation of CsrB in the cell would activate CsrA-repressed genes. In summary, the CsrB/C levels are a key determinant of CsrA activity in the cell. The 366-nucleotide CsrB RNA forms a globular ribonucleoprotein complex containing 18 CsrA subunits, where most binding sites are located in the loops of predicted hairpins (Fig. 6.37). The 245-nucleotide CsrC RNA possesses nine imperfect repeat sequences, primarily localized in the loops of predicted hairpins, serving most probably as CsrA binding elements. Other than these repeated elements, CsrB and CsrC sequences exhibit no striking similarity. The levels of all three components, CsrA, CsrB and CsrC, accumulate as the culture approaches the stationary phase of growth. Expression of *csrB* and *csrC* is activated via the BarA/UvrY two-component system.

The 6S Regulatory RNA Competes with Binding of RNA Polymerase to σ^{70}-specific Promoters

The 6S RNA specifically interacts with the holoenzyme carrying the housekeeping sigma factor, σ^{70}, makes direct contact with σ^{70} and decreases transcription from σ^{70}-dependent promoters. This small ncRNA accumulates to high levels as *E. coli* cells reach the stationary phase of growth when the alternative sigma factor, σ^{S}, is needed for maintaining the stationary phase to survive in nutrient-poor conditions. Blocking of σ^{70}-dependent transcription by 6S RNA may be one mechanism to allow σ^{S} to be an effective regulator during the stationary phase. In summary, the 6S RNA structure seems to mimic DNA at the promoter, competing with binding of RNA polymerase to σ^{70}-specific promoters.

An intriguing question with respect to studies of small ncRNA is: why use a regulatory RNA in preference to a protein? What are the benefits of the proposed regulatory mechanisms? A commonly used argument is *economy*: because these RNAs are small and untranslated, the energetic cost of their synthesis is much lower than for a protein. Thus, a small RNA should be more cost-effective than a protein. Chromosomally encoded riboregulators are stable RNAs with halflives usually exceeding ~15 min. Even transposase-encoded antisense RNAs such as RNA-OUT are moderately stable. In contrast, almost all copy number regulator RNAs as well as the antisense RNAs that control post-segregational killing are unstable, with halflives of typically around 1 min. Hence, it is conceivable that economy is only part of the answer. The other part could be design for the particular role. This implies that riboregulators whose synthesis is induced when required, e.g., during oxidative stress (OxyS), low temperature (DsrA) or various other stress conditions (MicF), should be stable throughout the stress response.

S. Gottesman **2005**, Micros for microbes: noncoding regulatory RNAs in bacteria, *Trends Genet.* 21, 399–404.

K.M. Wassarman **2002**, Small RNAs in bacteria: diverse regulators of gene expression in response to environmental changes, *Cell* 109, 141–144.

6.3.3 Riboswitches

RNA is known to possess many different functions in the cell, ranging from acting as a passive mRNA to actively controling the expression pattern of various genes as functional or regulatory RNA. Functional RNAs can either be part of the mRNA they control or be transcribed independently. In the first case, they act in *cis* and can control just the RNA molecule they are part of, whereas in the second case they act in *trans* and are able to interact with several target RNAs. *cis*-acting regulatory RNAs are located in the 5′UTRs of bacterial mRNAs, are designated *riboswitches* and act as metabolite sensors. In most cases, they turn off gene expression upon binding of the metabolite, but two cases have been described so far where binding triggers gene expression. Riboswitches act by one of three possible mechanisms: (a) premature transcription termination, (b) preventing translation initiation and (c) mRNA processing by acting as a ribozyme called *aptazyme*.

Temperature Sensing

The prototype for temperature sensing is the *E. coli rpoH* transcript. This gene codes for the alternative sigma factor, σ^{32}, which plays an important role in transcription for many heat-shock genes (see Section 9.2). At low temperatures, translation of the *rpoH* mRNA is largely repressed, while after a sudden increase in

temperature, translation occurs at a high level. Two DNA segments near the 5′ end form a complex secondary structure constituting the thermosensor. The presence of this secondary structure largely prevents the ribosomes from binding to the Shine–Dalgarno sequence. After a heat shock, the secondary structure is melted and binding of the 30S ribosomes to the mRNA allows high-level translation. A comparable mechanism was found to control expression of some heat-shock genes in rhizobia. The 5′ end of these heat-shock genes contains an untranslated region called ROSE (for *r*epression *o*f heat *s*hock genes *e*xpression) the length of which varies over 70–120 nucleotides. As described for the 5′ end of the *rpoH* mRNA, these transcripts adopt extended secondary structures sequestering the Shine–Dalgarno sequence and the AUG start codon, preventing access of the 30S ribosomal subunit. Here, too, higher temperatures destroy the secondary structure, allowing the ribosomes to initiate translation. There are additional mRNAs encoded by heat-shock genes that can adopt heat-sensitive secondary structures, including the transcript of the *dnaK-dnaJ* operon of *H. ducreyi* and the *hsp18* mRNA of *S. albus*. In addition, this concept is also used by pathogenic bacteria to express virulence genes only when they have entered the body of a warm-blooded host. All these examples suggest that temperature sensing by mRNA molecules seems to be widespread among eubacteria.

Vitamin Sensing

The presence of three different vitamins can be sensed by the 5′ ends of transcripts coding for the respective vitamin biosynthesis or uptake genes. One example is the *cob* operon of *S. typhimurium* responsible for the biosynthesis of cobalamin (vitamin B_{12}). While translation of the transcripts of the genes involved in the biosynthesis of this vitamin and its uptake are repressed by B_{12} itself, these mRNAs are efficiently translated under conditions of B_{12} deficiency. In the presence of B_{12}, one of several RNA hairpins present in the leader sequence sequesters the Shine–Dalgarno sequence. If B_{12} is absent, another hairpin, the so-called translational enhancer, helps to maintain the SD sequence in an accessible conformation through long-distance interaction. The binding of adenosyl cobalamin, a precursor of B_{12}, to the mRNA seems to prevent the translational enhancer from interacting with the SD sequence.

The second example is the thiamin biosynthesis operon of *R. etli*. Thiamin, also known as vitamin B_1, is a cofactor for many important enzymes and therefore essential for growth. Bacteria have genes for all the enzymes necessary to synthesize thiamin, but if adequate amounts are present in the environment, they will use those rather than producing their own. Synthesis of thiamin biosynthetic enzymes is prevented by feedback inhibition, where thiamin represses expression of the genes needed for its own synthesis. This occurs by direct interaction of thiamin with the mRNA coding for the enzymes involved in the biosynthsis of B_1. This mechanism has been studied extensively in *R. etli*, where the transcript contains a 211-bp untranslated leader that can adopt three different hairpin struc-

tures. The 5′ proximal one is called the *thi*-box, the second one sequesters the SD sequence and the AUG start codon, while the third hairpin acts as a transcriptional attenuator causing dissociation of the RNA polymerase. The *thi*-box contains the sequence AGGGA serving as a ribosome binding site. Transcription from the promoter preceding the whole operon is constitutive. In the absence of thiamin, the *thi*-box will not fold, but the 30S subunit can bind to the AGGGA sequence and is passed to the SD sequence, where translation starts. The translating ribosomes impede formation of the attenuator structure, allowing continuation of transcription. In contrast, in the presence of thiamin, the *thi*-box folds into the hairpin structure, preventing interaction with ribosomes. This leads to the formation of the two other hairpin structures, causing attenuation. All available data suggest that TPP itself is involved in stabilization of the *thi*-box structure. This *thi*-box is highly conserved upstream of the thiamin biosynthesis and uptake genes of both Gram-negative and Gram-positive eubacteria.

The third vitamin that is sensed by mRNA is riboflavin. In *B. subtilis*, the riboflavin biosynthetic operon is preceded by an untranslated leader of about 300 nucleotides which contains an evolutionary conserved sequence. This sequence can fold into two alternative structures, the anti-anti-terminator and the anti-terminator, where both structures have virtually equal stability. The anti-anti-terminator structure folds only in the presence of FMN/FAD.

Why might mRNA molecules act as a sensor? In both cases, secondary structures are involved which influence the translation efficiency. In the first case, no additional molecule is involved and just the melting of the secondary structure enhances translation. In the second case, an effector molecule is involved which is assumed to stabilize a secondary structure and thereby prevent translation. This behavior is reminiscent of a group of artificial RNAs called *aptamers* which bind with high affinity and specificity to a wide range of targets, including small molecules such as amino acids, nucleotides, antibiotics and enzyme cofactors. In general, aptamers for small molecules contain bulges and internal loops that can adopt specific binding pockets on interaction with the ligand. In the case of the *thi*-box, this RNA structure contains such an internal loop that can participate in a binding pocket for thiamin. Therefore, it is easy to imagine that natural mRNAs might contain specific binding sites for some small effectors that could be used in regulating gene expression. If it can be shown that thiamin, cobalamin and riboflavin are effectors that act directly on the mRNA, one can ask why such a mechanism is not common. In summary, one might argue that sensory mRNAs are relicts with ancient origins which have been replaced in most cases by regulation mechanisms involving proteins.

Sensing of Other Metabolites

Many more examples of metabolite-controlled riboswitches have been reported. For instance, one of the recently identified riboswitches responds to the coenzyme *S*-adenosylmethionine (SAM), which binds its target with a dissociation constant of ~4 nM. This riboswitch can discriminate ~100-fold against *S*-adenosylhomocys-

teine that differs from SAM by a single methyl group and an associated positive charge. Another class of riboswitches selectively recognizes guanine and becomes saturated at concentrations as low as 5 nM. The last example to be mentioned here is the 5′UTR of the *lysC* gene of *B. subtilis* that serves as a lysine-responsive riboswitch. The ligand binding domain of the riboswitch binds L-lysine with an apparent dissociation constant of ~1 μM and discriminates against closely related analogs such as D-lysine and ornithine.

Induction of Glycine-degrading Operons

There is one example so far where binding of a metabolite induces relief of transcription attenuation. This riboswitch is termed *gcvT* and consists of the two aptamers I and II (Fig. 6.38). The *gcvT* motif has been identified in several bacterial species and analyzed in *B. subtilis* and *V. cholerae*. The biological function of these operons is to endow the cells with the potential to catalyze the initial recations for the use of glycine as an energy source. If the glycine concentration within the cells is low, the amino acid will not bind to the two aptamers and a transcription terminator is formed, as shown for the *B. subtilis* operon in Fig. 6.37. If, on the contrary, the concentration is high (10 mM in the experiment), binding of glycine molecules to the two aptamers stabilizes their secondary structures, preventing formation of the transcription terminator, and the two structural genes *gcvPA* and *gcvPB* are expressed.

A.G. Vitreschak, et al. **2004**, Riboswitches: the oldest mechanism for the regulation of gene expression? *Trends Genet.* 20, 44–50.

W.C. Winkler, R.R. Breaker **2005**, Regulation of bacterial gene expression by riboswitches. *Annu. Rev. Microbiol.* 59, 487–517.

6.3.4 mRNA Decay

Besides the efficiency of transcription of DNA in mRNA, the stability of the mRNA also influences the level of gene expression. In exponentially growing *E. coli* cells, mRNA halflives of different transcripts range from less than 30 s to more than 15 min (the *ompA* transcript serves as a paradigm), disclosing an enormous potential to adjust the intracellular levels of transcripts by means of differences in stability. The halflife of the bulk of *E. coli* mRNA has been determined to be 2.4 min at 37 °C. mRNA halflives can be defined by two principal values: *chemical* and *functional*. The chemical halflife measures the decay of transcripts after transcription has been blocked by the addition, for instance, of rifampicin. The functional halflife measures the inactivation of a transcript by monitoring protein synthesis after transcription has been blocked. Here, a single endonucleolytic cut within the coding region immediately destroys the transcript. The following *cis*- and *trans*-acting factors have been described as being involved in influencing the halflife of mRNA molecules:

- determinants located at the 5′ end
- determinants located at the 3′ end
- exonucleases
- endonucleases
- ribosomes
- polyadenylation.

The length of *E. coli* mRNAs can vary from a few hundred, for monocistronic transcripts, to many thousand nucleotides. It was anticipated that longer transcripts are more susceptible to degradation than shorter mRNAs, due to the expected

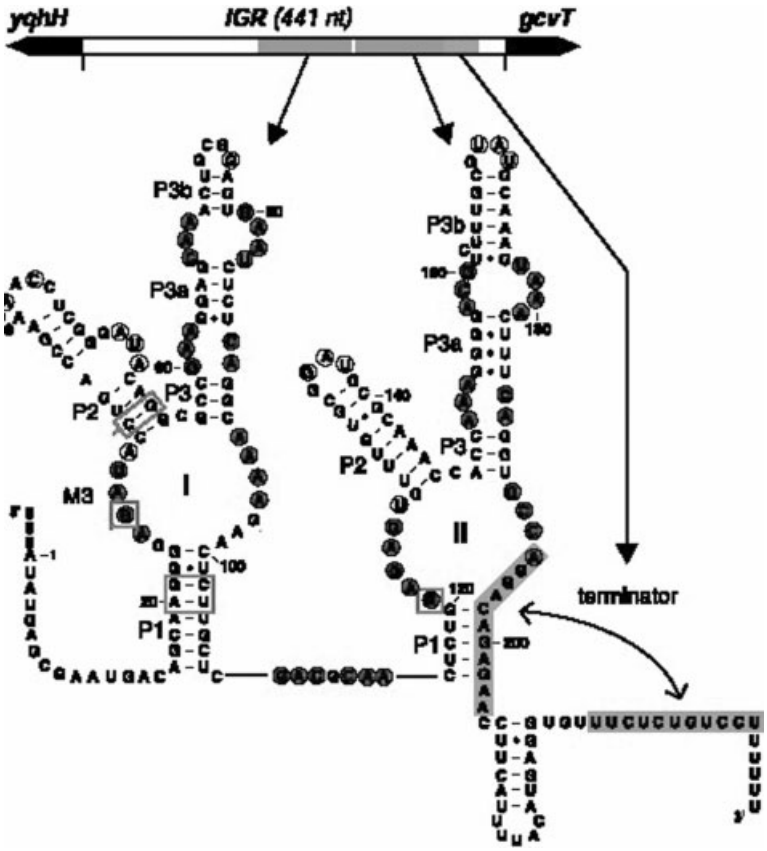


Fig. 6.38 The *B. subtilis gcvT* riboswitch. The intergenic region between *yqhH* and the *gcvPA-gcvPB* operon designated *gcvT* acts as a riboswitch and contains the two glycine sensing aptamers I and II. If the glycine concentration in the cell is low, the two aptamers remain unoccupied favoring formation of the terminator structure resulting in transcription attenuation. If the glycine concentration is high, it will bind to the two aptamers preventing formation of the transcription terminator, and the two structural genes will be expressed. M. Mandal, et al. **2004**, *Science* 306, 275; Fig. 4, modified. (This figure also appears with the color plates.)

number of cleavage sites for endoribonucleases. This turned out to be wrong. While the stability of the monocistronic *lpp* transcript (325 nucleotides) has been measured to be 15 min, that of the *trxA* transcript (493 nucleotides) has been determined as 2–3 min and that of *trmD* (2100 nucleotides) at 3 min. There are more examples, all proving that the length of a given transcript does not correlate with its halflife. Furthermore, transcript degradation in *E. coli* appears to operate at the level of the cistron, not the full-length message. This has been shown for the polycistronic *trp* transcript where the mRNA levels of cistrons vary considerably, as does the synthesis of the corresponding enzymes.

Determinants Located at the 5′ End of Transcripts

The 5′ determinants of relatively stable mRNAs have been identified by their stability to confer to unstable transcripts by swapping of the corresponding region. The 5′ stabilizer of the *E. coli ompA* and the *S. aureus ermC* transcripts are among the best characterized 5′ determinants. The secondary structure of the *ompA* 5′UTR is composed of two stem-loop structures located between the 5′ terminus and the ribosome binding site, but only the 5′ proximal hairpin structure is essential for the relatively long halflife of the transcript. Extending the 5′ with five unbasepaired nucleotides completely abolishes the stabilizing effect of the terminal hairpin structure, suggesting that a single-stranded 5′ extremity allows ribonuclease better access to mRNA. RNase E has been identified as the enzyme primarily interacting with the 5′ ends of mRNA.

Determinants Located at the 3′ End of Transcripts

Stable hairpins located at the 3′ end of many mRNA, including Rho-independent terminators or REP elements, protect the upstream mRNA from degradation by the two known exonucleases. Transcripts without such structures are extremely vulnerable to rapid degradation up to the next stem-loop structure. A single stem-loop structure normally seems to exert the same stability to a transcript as a series of such structures at the 3′ end.

Internal Stem-loop Structures

Internal hairpin structures can be often found within polycistronic transcripts and are able to separate polycistronic transcripts into cistron segments with differential stabilities. Endoribonucleolytic initiation of mRNA degradation occurring downstream of an intercistronic stem-loop structure leads to subsequent 3′→5′ exoribonucleolytic degradation, up to the secondary structure. If the cleavage step takes place upstream of the hairpin structure, it promotes rapid degradation of the upstream segment. Alternatively, the internal stem-loop itself can serve as a recognition site for endoribonucleolytic cleavage. One example is the tricistronic transcript *metY-nusA-infB*, which is processed by RNase III at the internal hairpin structure between *metY* (coding for a tRNA) and *nusA*. This cleavage separates the

structural tRNA from the informational mRNA and initiates the rapid degradation of the 5′ extremity of the downstream mRNA. In summary, internal stem-loop structures exert an important regulatory function by directing differential degradation of adjacent segments of transcripts and thereby contribute to differential expression of genes within an operon.

Role of Exoribonucleases

Two 3′→5′ exoribonucleases have been described which involve the degradation of mRNA, namely RNase II (*rnb*) and polynucleotide phosphorylase (PNPase; *pnp*), where the latter enzyme seems to be the primary exoribonuclease involved in the degradation of mRNAs. RNase II, a monomer of 64 kDa, is a highly processive and hydrolytic enzyme converting RNA into monophosphates and is strongly impeded by secondary structures. PNPase, present in a pentameric form of 320–360 kDa, is a phosphorolytic enzyme which converts RNA into nucleotide diphosphates in a reversible reaction. Both enzymes are different at their primary structure and are functionally redundant. *E. coli* can tolerate a knockout of either gene without exhibiting a phenotype, but a double knockout is not viable. Both exonucleases require free single-strand 3′ ends for binding to their substrates. Thus, the ternary elongation complex containing the DNA template and the RNA polymerase protects the nascent transcript from the exonucleases. Released transcripts are protected by the Rho-independent termination structure which locks the 3′ end into a stable, double-stranded structure. The short oligo(U) stretch at the 3′ end is too short for exonuclease binding. A third type of barrier against the exonucleases involves REP sequences (see Section 2.3.2.1), where *E. coli* contains about 600 REPs located in intercistronic regions in single or multiple copies. After transcription, they form stable stem-loop structures. Many additional 3′→5′ riboexonucleases, including RNases R (*rnr*), BN, PH, D and T, have been described but these seem to be exclusively involved in the final maturation of the 3′ end tRNAs. Both the *pnp* and *rnb* genes are cold-inducible (see Section 9.3). No 5′→3′ exonuclease has ever been discovered. There is one additional exoribonuclease involved in mRNA decay. This enzyme, called oligoribonuclease (*orn*), is essential for cell viability. Since neither PNPase nor RNase II can completely degrade an oligoribonucleotide, leaving 4- to 7-mers as terminal digestion products, these short oligonucleotides accumulate in an *orn* mutant.

Role of Endoribonucleases

E. coli is known to contain at least five endoribonucleases which all operate under tight cellular regulation and are involved in the modification, maturation and degradation of ribonucleic acids. Those enzymes known to participate in maturation of stable RNAs from precursor RNAs are designated *processing RNases* and those involved in degradation of transcripts *degrading RNases*. *RNase III*, like other processing endoribonucleases, generates 5′ P and 3′ OH, while the degradative RNases generate 5′ OH and 3′ P ends. This enzyme was first discovered as a pro-

tein that cleaves double-stranded RNA. *In vivo*, RNase III specifically degrades stem-loop structures, particularly those in intercistronic regions (see above). Furthermore, it plays a major role in processing the 30S rRNA precursor. Deletion of the RNase III structural gene (*rnc*) does not lead to any significant change in the degradation of total pulse-labeled RNA and therefore plays a minor role in mRNA decay. The 118-kDa *RNase E* encoded by the *rne* gene is involved in the processing of a large number of nontranslated *E. coli* transcripts, such as 9S rRNA, into the 5S form, the processing of the 5′ end of the 16S rRNA, the maturation of the RNA subunit of RNase P and the processing of tRNAs. In addition, RNase E is implicated in the decay of total pulse-labeled RNA, as well as a number of specific transcripts. The RNase E protein consists of 1061 amino acids, comprising at least three distinct domains. The N-terminal 500 amino acids (aa) encode the catalytic activity of the protein and are responsible for the preferential partitioning of the enzyme near the cell membrane. The middle domain (aa ~597 to 684) is the so-called arginine-rich RNA binding site that has been shown to bind RNA *in vitro*. The C-terminal third of RNase E (aa ~734 to 1061) functions as a scaffold for the assembly of a multiprotein complex called the *degradosome* (Fig. 6.39). The RNase E scaffold region contains sites for the binding of the major [PNPase, the RhlB ATP-dependent RNA DEAD-box helicase (*in vitro*, this helicase can be replaced by the helicases RhlE or CsdA), the glycolytic enzyme enolase] and minor (GroEL, DnaK, polyphosphate kinase) degradosome components. Enolase within the degradosome plays a crucial role in the regulation of *ptsG* mRNA stability in response to metabolic stress. RNase E prefers single-stranded regions that are typically A-U rich (see below). In addition, RNase E is a 5′ end-dependent endoribonuclease that prefers substrates with monophosphorylated 5′ ends to triphosphate termini. Recently, a 17.4-kDa protein named RraA (*r*egulator of *r*ibonuclease *a*ctivity *A*) has been described. This protein binds to RNase E and inhibits RNase E endonucleolytic cleavages without altering cleavage site specificity or interacting detectably with substrate RNAs. What might be the putative function of RraA? This protein might facilitate rapid alterations in RNA decay and/or processing in response to specific environmental stimuli. In an *rraA* null strain, the number of trancripts affected is relatively small and the cells exhibit normal growth. The lack of a strong phenotype of an *rraA* null mutation may be due to the presence of one or more additional proteins capable of modulating the

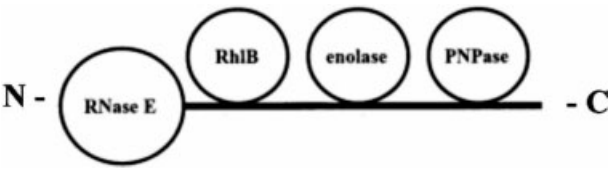


Fig. 6.39 Schematic structure of the *E. coli* RNA degradosome. The circle marked with RNase E contains the catalytic activity of the protein and the RNA binding site. The remaining part serves as a scaffold for the three proteins shown, and the four-protein complex is termed RNA degradosome. V. Khemici, A.J. Carpousis **2004**, *Mol. Microbiol.* 51, 777–790; Fig. 6A, modified.

catalytic function of RNase E. Indeed, such a protein has been found and designated RraB. Homologs of RNase E or its close relative, *RNase G* (*cafA*), a second 5′ end-dependent endoribonulease, have been identified in or their existence inferred from the DNA sequence analysis of more than 50 bacterial species. *RNase I* (*rna*) is a relatively nonspecific endoribonuclease that is found primarily in the periplasmic space. This enzyme could be involved in the degradation of RNA molecules taken up by the cells. A modified version of this enzyme, *RNase I*, is present within the cytoplasm, but a *rna* knockout has no phenotype. Furthermore, a mutant form of RNase I, called *RNase G*, has been described in only one *E. coli* strain.

Ribosomes

Ribosomes bound to the Shine–Dalgarno sequence can protect the 5′ end of transcripts from binding to the RNA degradosome. Furthermore, ribosomes translating part of the message at a high rate can reduce endonucleolytic attacks. In the case of the *malEFG* operon, it has been suggested that ribosomes translating *malE* block the distal binding of RNase E, shifting the enzyme to a more poorly translated site somewhere in the distal *malF* transcript. Another well documented example is the *cryIIIA* gene of *Bacillus thuringiensis*. This gene codes for a toxin belonging to the group of so-called crystal proteins produced during the stationary/sporulation phase of growth. Crystal protein synthesis is massive (in some subspecies the crystal is 20–30% of the dry weight of the sporulated cells). The transcript coding for this protein has an halflife of about 10 min, which is mainly based on a 5′ mRNA stabilizer. The determinant of the stabilizer is an SD sequence located in the 5′UTR and not involved in translation initiation (Fig. 6.40). Mutations prediced to weaken the putative interaction between the SD sequence and the 3′ end of the 16S rRNA lead to reduced mRNA stability.

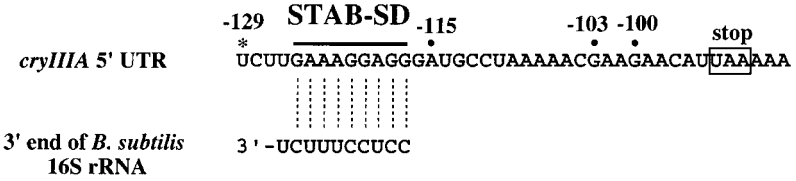


Fig. 6.40 Stabilizer Shine–Dalgarno sequence upstream the *cryIIIA* gene of *B. thuringiensis*. The *cryIIIA* gene of *B. thuringiensis* is is preceded by strong Shine–Dalgarno sequence termed STAB-SD, to which ribosomes bind without initiating translation though it is followed by a very short open reading frame, but the AUG is to close to the Shine–Dalgarno sequence. H. Agaisse, D. Lereclus **1996**, *Mol. Microbiol.* 20, 633–643; Fig. 5.

Role of Polyadenylation

Polyadenylation of the 3′ ends of RNAs, once thought to occur exclusively in eukaryotic cells, has been shown definitely to occur in bacterial systems as well and

is now considered in *E. coli* as a general RNA quality control mechanism. In bacteria, the level of polyadenylation is a key determinant of RNA stability. Thus, RNAs with poly(A) tails are targeted for degradation by the two major exonucleases, RNase II and PNPase, in bacterial cells. *E. coli* cells contain at least two enzymes capable of adding poly(A) tails to the 3′ ends of RNA molecules. The major polyadenylating enzyme, designated poly(A) polymerase I (PAP I), was originally identified by virtue of its role in regulating plasmid copy number and is the product of the *pcnB* (for *p*lasmid *c*opy *n*umber *B*) gene. Mutants of *E. coli* lacking PAP I still retain the ability to polyadenylate RNAs, indicating that there is at least one other polyadenylating enzyme in those cells. A recent report provides evidence for the function of polynucleotide phosphorylase as the second poly(A) polymerase in *E. coli*. Interestingly and unlike the situation in eukaryotes, both ribosomal and messenger RNAs in *E. coli* are polyadenylated.

Sequence analysis of poly(A) tails has detected the presence of occasional C, U and G residues. While PAP I has the highest preference for ATP as substrate, it is also capable of polymerizing very small amounts of UTP, CTP and GTP into mixed polymers, in the order: ATP > CTP > UTP > GTP. At any given time, only a small percentage (1–2%) of the total RNA population is polyadenylated. But increasing the level of poly(A) polymerase can lead to quantitative polyadenylation of a particular transcript. Polyadenylation can occur not only on mature mRNAs, but also on fragments resulting from endo- and exonucleolytic degradation and even on stable RNAs or their precursors. In fact, many accessible 3′ ends seem to be a potential target for polyadenylation in *E. coli*. In addition, poly(A) tail length in wild-type cells is relatively short (10–40 nucleotides) compared with 70–90 nucleotides in yeast and over 200 nucleotides in vertebrates. In contrast to yeast, the addition of poly(A) tails to *E. coli* transcripts appears to target them for decay (see below).

mRNA Decay Models

Two different models to explain decay of mRNA have been suggested, one relying on nonspecific endoribonucleases and the second employing RNases E and G. In the first model, an enzyme such as RNase M, a mutant form of RNase I (see above), initiates mRNA degradation (Fig. 6.41A). The mRNA fragments are then degraded by exonucleases to oligoribonucleotides. If stem-loop structures are close to the 3′ ends, PAP adds a poly(A) tail to facilitate binding and RNA helicases unwind double-stranded region. In the second model, RNase E (or RNase G) not associated with the RNA degradosome binds at the 5′ terminus and initiates degradation by cleaving at a distance (Fig. 6.41B). Then, the enzyme moves in the 5′→3′ direction, generating mRNA fragments that are completely degraded as described in model one. One problem is posed by the 5′ end of the transcripts which contain three phosphate groups. RNase E prefers a 5′ PO_4 to a 5′ triphosphate, which is found at the primary 5′ terminus of each transcript. Perhaps, there is an enzyme which can remove a diphosphate from the 5′ end to allow binding of RNase E.

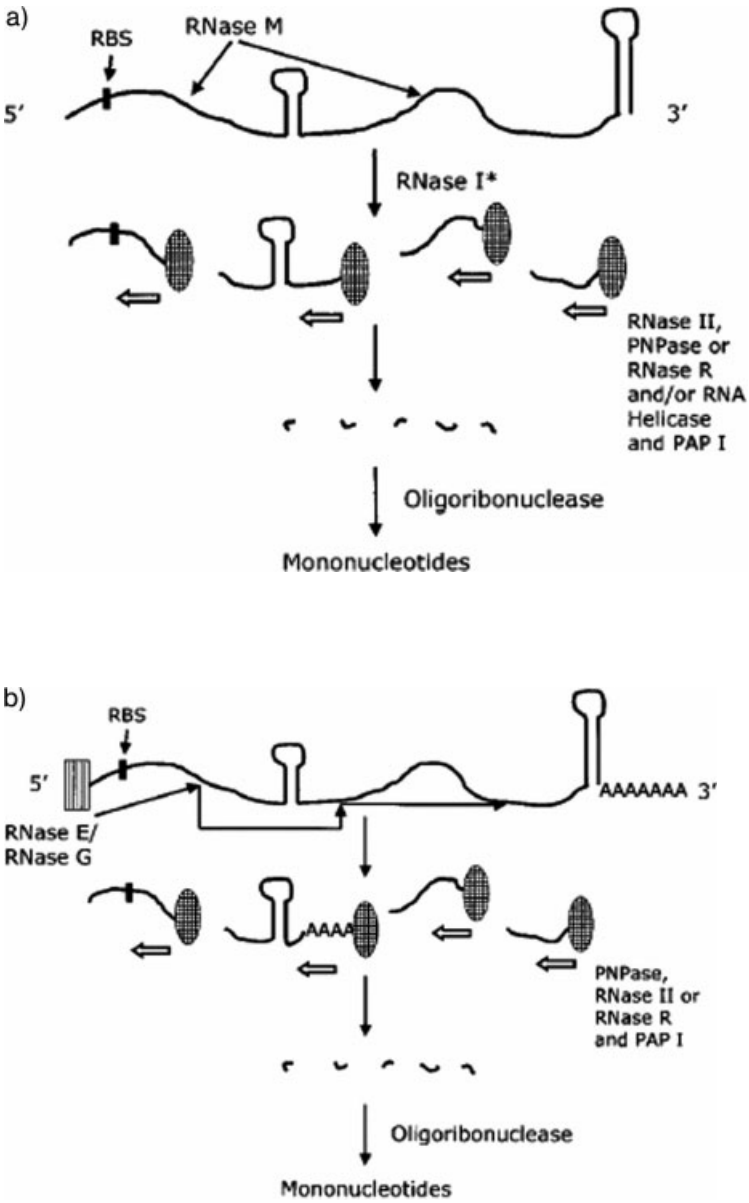


Fig. 6.41 mRNA degradation models. (A) An unspecific endoribonuclease such as RNase M cleaves at one or more internal sites and the resulting fragments are then degraded from their 3′ ends by an exonuclease. If the free 3′ ends downstream of stem-loop structures are too short to allow binding of an exonuclease, they are elongated by an poly(A) tail using PAP. (B) RNase E or RNase G bind to the 5′ end and cleave at a distance followed by by a movement in the 5′→3′ direction and additional cleavage. Degradation of the RNA fragments occurs as described under A. S.R. Kushner **2002**, *J. Bacteriol.* 184, 4658–4665; Figs. 2, 3.

It is assumed that RNase E, and possibly RNase G, serves as the enzyme that initiates mRNA degradation. One of these two enzymes starts mRNA decay by binding to the 5′ terminus and then cleaving at a distance. Since the enzyme prefers a 5′ PO_4 to a 5′ triphosphate, the initial cleavage reaction would be rate-limiting. This model could explain why some 5′ untranslated regions, such as the highly structured *ompA* leader, impart enhanced stability to a variety of transcripts. The shorter oligoribonucleotides produced by RNase E can be degraded by PNPase. Since PNPase is inhibited by secondary structures, the presence of the RhlB RNA helicase can promote potential inhibitory structures. Alternatively, or in addition, the poly(A) tails present at the 3′ termini of Rho-independent transcription terminators would provide a binding site for PNPase.

A.J. Carpousis, et al. **1999**, mRNA degradation: a tale of poly(A) and multiprotein machines, *Trends Genet.* 15, 24–28.

D. Kennell **2002**, Processing endoribonucleases and mRNA degradation in bacteria, *J. Bacteriol.* 134, 4545–4557.

S.R. Kushner **2002**, mRNA decay in *Escherichia coli* comes of age, *J. Bacteriol.* 184, 4658–4665.

Degradation of Stable RNA

Although tRNA and rRNA, accounting for ~98% of the cellular RNA, are quite stable during exponential growth, under certain physiological conditions these RNA molecules can be extensively degraded. Degradation of stable RNA is associated with starvation and stationary phases, both of which have in common the absence of growth. Conditions of starvation can be induced by the depletion of any one of a number of nutrients, including nitrogen, carbon, phosphate or even Mg^{2+}. All these conditions lead to a dramatic loss of RNA, which is confined largely to rRNA. The degradation of ribosomes during starvation can be rapid and quite extensive, amounting to >96% in some studies. In *Salmonella* strains, rRNA degradation is taken to extremes. More than 90% of 23S rRNA and ~50% of the 16S rRNA are degraded when cells reach the stationary phase. The nonspecific endoribonuclease *RNase I* is responsible in many of these situations for RNA degradation. RNase I is thought to be present largely in the periplasmic space, but membrane damage would allow entry of active enzyme into the cytoplasm. In addition, loss of Mg^{2+}, an inhibitor of RNase I, and the more exposed rRNA would lead to extensive rRNA degradation. Degradation of stable RNA can be induced by the F-factor-encoded gene *srnB*, whose 12-kDa protein increases membrane permeability and allows RNase I to enter the cytoplasm. Under normal conditions, the *srnB* gene is turned off directly or indirectly by the chromosomal gene *srnA*. Mutations of *srnA* or overexpression of *srnB* promote stable RNA degradation.

Another mechanism of degradation of stable RNAs involves those molecules which have a defect. Synthesis, maturation and assembly of RNA molecules into ribonucleoprotein particles are carried out in cells with a high degree of accuracy. However, these complex processes are not perfect; and these nonfunctional RNAs have the potential to interfere with the function of their normal counterparts.

Therefore, precursors of stable RNAs that cannot be converted to their mature forms become polyadenylated. Because polyadenylation is a signal for RNA degradation, these defective precursors are removed. Therefore, *E. coli* has a quality control mechanism for the specific elimination of defective stable RNA precursors.

6.3.5 Introns

Introns or intervening sequences interrupt genes and are removed from the precursor RNA by a process called *RNA splicing*. In eubacteria, two classes of introns have been described, *group I* and *group II* introns, where group I are typically found both in bacteriophage mRNAs coding for proteins and in tRNAs, and group II are found in mRNA only. Group II introns are rare and occur in mobile elements such as conjugative plasmids and transposons. Both intron groups are distinguished from one another in their sequence and structure as well as in their splicing mechanism. They are capable of splicing themselves out of RNA without the help of proteins. Therefore, intron RNAs are enzymes and RNA enzymes are called *ribozymes*. Other known bacterial ribozymes are RNase P and the 23S rRNA peptidyltransferase. Some of the group I and II introns are mobile; and they are able to spread through a population of homologous intronless alleles by a process called *retrohoming*.

Group I Introns

The first bacterial intron, a self-splicing group I intron, was found in 1984 to interrupt the thymidylate synthase (*td*) gene of phage T4. The next two group I introns were detected in the aerobic (*nrdB*) and anaerobic (*nrdD*) ribonucleotide reductases of the same phage. In bacteria, group I introns never interrupt protein-coding genes, but in most cases one of four different tRNA genes. There is also one example of an interrupted bacterial rRNA gene (23S rDNA). The controverse is true for phage-encoded group I introns that are always found within protein-coding genes. Homing endonucleases are typically encoded within phage introns and not in chromosomal introns, with two exceptions. Group I introns exhibit minimal primary sequence conservation but have conserved secondary and tertiary structures that are required for splicing. The exogenous guanosine cofactor that promotes splicing initiates a series of transesterifiction reactions, resulting in intron splicing and exon ligation.

Group II Introns

Group II introns are retroelements of a highly structured catalytic RNA and a multifunctional, *i*ntron-*e*ncoded *p*rotein (IEP), which possesses three biochemical activities: RNA maturase that facilitates splicing, reverse transcriptase and DNA endonuclease. They occur in cyanobacteria, proteobacteria, Gram-positive bacteria, archaea and the organelles of eukaryotes. They are *ribozymes* that catalyze the

splicing of their flanking exons and are hypothesized to be the progenitors of eukaryotic spliceosomal introns. About 25% of the completely sequenced microbial genomes contain one or more introns (either full-length or fragmented). Group II introns are often inserted in intergenic regions, and when located inside genes, they are rarely found within highly conserved or housekeeping genes. Usually, these introns are located on mobile DNA elements, such as plasmids, IS elements, transposons or pathogenicity islands, which could account for their spread among bacteria. Some group II introns are also mobile retroelements. They move site-specifically, either to homologous intronless genes via an RNA intermediate in a process called *retrohoming*, or into novel (ectopic) genomic locations (but at a much lower frequency) which is termed *retrotransposition*. The homing process is highly specific and occurs at target regions spanning ~30 bp around the insertion point. Selection of the splice sites is determined by base pairings between three motifs in the intron RNA (exon-binding sequences EBS1, EBS2, EBS3) and the complementary sequences in the flanking exons (intron-binding sequences IBS1, IBS2, IBS3, respectively) and these pairings are required for both splicing and homing.

Splicing of group II introns occurs via two sequential transesterification reactions (Fig. 6.42). In the first, nucleophilic attack at the 5′ splice site by the 2′ OH of a bulged A-residue results in cleavage of the 5′ splice site, which is coupled to the formation of a lariat intermediate. The second nucleophilic attack occurs at the 3′ splice site by the 3′ OH of the cleaved 5′ exon and results in exon ligation and release of the intron lariat. Intron splicing is promoted by the maturase function of the IEP. The maturase is encoded by a self-contained open reading frame with an independent promoter.

The best studied bacterial group II intron is the L1.LtrB intron of *L. lactis* ML3, which is inserted in the conjugative relaxase gene essential for the transfer of the plasmid pRS01. This mobile retroelement is capable of both efficient retrohoming into its cognate intronless gene and low-frequency *retrotransposition* to ectopic sites. Both processes require a complex that contains the IEP LtrA and the excised intron RNA, with both being used for DNA target site recognition.

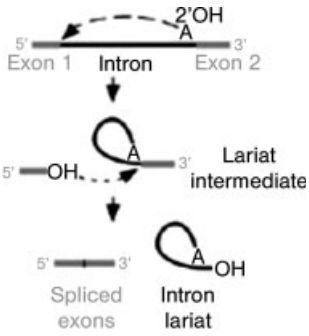


Fig. 6.42 The group II intron splicing mechanism. Intron splicing occurs through two consecutive transesterification reactions. First, the 2′ OH of a bulged A-residue carries out a nucleophilic attack at the 5′-splice site leading to the formation of a lariat intermediate (upper part). Second, a nucleophilic attack at the 3′-splice site by the 3′ OH of the cleaved 5′ exon (middle part) results in exon ligation and release of the intron lariat (lower part). A.M. Lambowitz, S. Zimmerly **2004**, *Annu. Rev. Genet.* 38, 1–35; Fig. 1.

Homing and Retrohoming of Introns

Homing and retrohoming is completely dependent on and initiated by intron-encoded proteins and relies on host-encoded proteins for completion. Group I introns carry out homing which is initiated by a DNA endonuclease termed *homing endonuclease* encoded within the intron itself. Based on conserved amino acid motifs and structural similarities, four families of endonucleases have been identified, wherein all endonucleases recognize lengthy DNA sequences (14–40 bp) that span the analogous intron insertion site in the intronless allele and normally introduce a double-strand break (Fig. 6.43). Host- and phage-encoded DNA repair proteins use the intron-containing allele as a template to repair the double-strand break and copy the intron into the intronless allele, leading to the formation of a Holliday structure which, depending on the resolution sites, leads to two different products as shown in Fig. 6.43. Phage T4 codes for two different repair pathways, DSBR (for *d*ouble-strand *b*reak *r*epair) and SDSA (for *s*ynthesis-*d*ependent *s*trand *a*nnealing; Fig. 6.43). While in DSBR, resolution of the Holliday junction results in both crossover and noncrossover products, in SDSA no Holliday junctions are found.

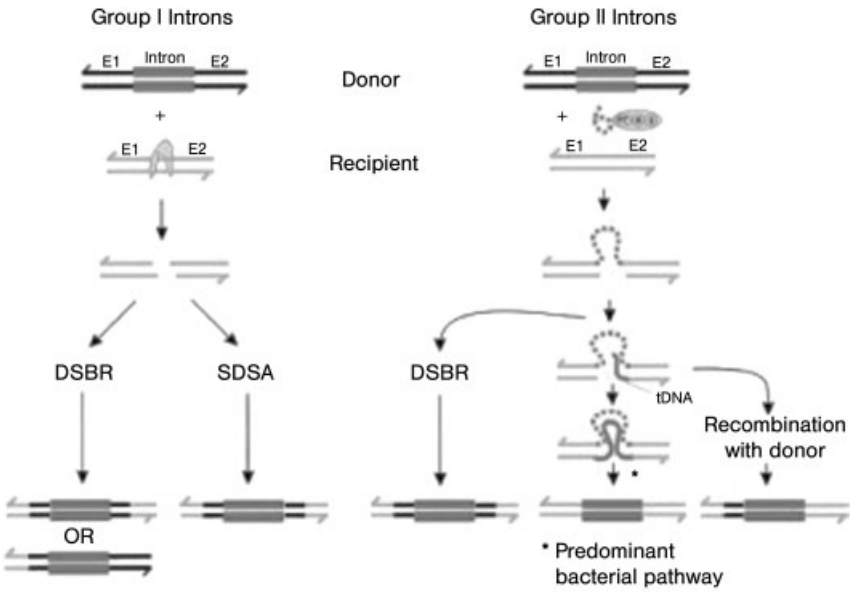


Fig. 6.43 Group I homing and group II intron retrohoming pathways. During group I homing, the intron-encoded homing endonuclease (shown in grey) catalyzes a double-strand break within the intronless allele which is normally repaired by *d*ouble-strand *b*reak *r*epair (DSBR) process involving host- and phage-encoded proteins. Phage T4 codes for two different repair pathways, DSBR and SDSA (synthesis-*d*ependent strand *a*nnealing). Group II retrohoming starts with an endonucleolytic break in the sense DNA strand catalyzed by the intron bound to the intron-encoded protein (shown in gray) and retrosplicing of the intron into the nicked DNA. See text for further details. D.R. Edgell, et al. **2000**, *J. Bacteriol.* 182, 5281; Fig. 1.

During group II retrohoming, the lariat-like intron RNA forms a complex with the *i*ntron-*e*ncoded *p*rotein (IEP) which binds to the target sequence of the intronless allele (Fig. 6.43). Next, the intron RNA cleaves the sense strand of the DNA by a reverse splicing reaction, resulting in the covalent insertion of the intron. Cleavage of the antisense strand is carried out by the endonuclease function of the IEP, and the 3′ end serves as a primer for cDNA synthesis by the reverse transcriptase function of the IEP, using the integrated intron as a template. In the bacterial pathway, the intron RNA undergoes a reverse splicing reaction and the resulting gap is filled by DNA. Yeast mitochondrial group II intron retrohoming proceeds through a recombination reaction with the donor DNA.

A.M. Lambowitz, S. Zimmerly **2004**, Mobile group II introns, *Annu. Rev. Genet.* 38, 1–35.

6.3.6
RNA-binding Proteins

Three classes of RNA-binding proteins have been described:

- translational regulators
- RNA chaperones
- RNA helicases.

Translational Regulators

Translational regulators are proteins which bind to mRNAs and act either as repressor or activator proteins. Examples are discussed under Section 6.2.1.

RNA Chaperones

RNA chaperones are a class of RNA binding proteins that, unlike helicases, have the ability to remodel structured RNAs in an ATP-independent fashion. Chaperone functions include the resolution of kinetically trapped, or misfolded secondary or tertiary RNA structures, as well as RNA annealing, helix destabilization and strand exchange activities. Here, we will consider two different RNA chaperone proteins, the global regulator Hfq and the FinO protein which plays an important role in conjugative transfer of F family plasmids.

The basic, heat-stable 11.2-kDa Hfq protein was first identified as a *h*ost *f*actor for replication of *Q*β RNA bacteriophage, where it functions to mediate structural changes of this RNA phage. Hfq proteins belong to the eukaryotic and archaeal family of Sm and Sm-like proteins and have been identified in several bacteria. They are strikingly conserved and highly abundant; and it has been estimated that there are 50 000–60 000 copies per rapidly growing *E. coli* cell, where the vast majority of Hfq is present in association with the translational machinery, but a minor fraction seems to be associated with the nucleoid nonspecific binding of Hfq to DNA or to RNA transcripts associated with the nucleoid. Since the active form of Hfq is a homohexamer, there are about 10 000 oligomers per cell, corresponding to an

intracellular concentration of ~15 μM. During the transition from growth to stationary phase, the level of Hfq decreases gradually to about one-third of the log phase number. Hfq is a pleiotropically acting RNA-binding protein that is required for the degradation of some mRNA transcripts and the efficient translation of others. Hfq targets several mRNAs for degradation by binding to their poly(A) tails and stimulating poly(A) adenylation. It is also believed to interact with and destabilize the *ompA*, *mutS* and *miaA* mRNAs, potentially through effects on their secondary structures and/or accessibility to ribosomes. In addition, Hfq has been shown to interact with several small noncoding RNAs (ncRNAs), such as OxyS, DsrA, RprA and Spot42, and is required for RNA regulation of the σ^S gene by OxyS, DsrA and RprA. These interactions typically alter the stability of the target transcripts. Hfq promotes contacts between OxyS and Spot42 molecules and their target RNAs, suggesting that Hfq assists in bimolecular RNA–RNA interactions. Moreover, studies carried out to identify additional small RNAs found that Hfq interacts with over half of these RNA molecules (~9 RNA species).

Hfq is a global regulator of *E. coli* metabolism and disruption of the *hfq* gene can cause a pleiotropic phenotype. The broad impact of the protein appears to stem from its role in regulating the stability and/or translation of mRNAs from a number of regulatory genes such as the *rpoS* transcript. Mutational studies suggest that Hfq is involved in the processes that effect the secondary structure near the 5′ end of *rpoS* mRNA, alleviating an inhibition of ribosome access to a translation start region. Hfq has been shown to affect the *in vivo* stability of mRNAs expressed from *ompA*, *mutS*, *miaA* and *hfq* genes and to stimulate elongation of the poly(A) tail of the *rpsO* mRNA.

The full-length, 77-residue *S. aureus* Hfq forms a hexameric ring with a width of ~23 Å and a diameter of ~65 Å. The central hole has a diameter of ~12 Å at the smallest constriction. Each Hfq consists of a bent five-stranded antiparallel β-sheet capped by an N-terminal α-helix. The RNA is bound in a circular, unwound manner around the pore of the Hfq hexamer, where the nucleotides bind in six separate but linked binding pockets, which spiral around the pore. The architecture of the Hfq-RNA complex suggests two, not mutually exclusive, mechanisms by which Hfq might exert its function as a modulator of RNA–RNA interactions. In the first mechanism, binding of single-stranded RNA leads to unwinding in a circular manner. This greatly destabilizes surrounding RNA structures located several nucleotides on either side of the binding site, permitting new RNA–RNA interactions. In the second mechanism, the repetition of identical binding pockets on the Hfq hexamer implies that the binding surfae can accommodate more than just a single RNA target. This would allow simultaneous binding of two RNA strands, thereby enhancing interaction between the strands.

The second example deals with the FinO RNA chaperone. The F family of conjugative plasmids confers antibiotic resistance and virulence to a wide variety of enterobacteria. Conjugative transfer is repressed by the protein FinO, which facilitates interactions between the mRNA of the major F-plasmid transcriptional activator, *traJ*, and the 79-nt antisense RNA, *finP*. FinO acts as a RNA chaperone by destabilizing otherwise stable stem structures in *finP* and *traJ* mRNA required for

the ability of FinO to promote strand exchange and duplexing between internal secondary structures within *finP* and *traJ* RNAs that would otherwise act as a kinetic trap to sense–antisense pairing. It has been suggested that FinO may use its own RNA binding free energy to destabilize a limited number of intramolecular base pairs in the stem-loop. In addition, FinO protects *finP* against endonucleolytic degradation.

RNA Helicases

Helicases are enzymes that use the energy of ATP hydrolysis to disrupt nucleic acid secondary structures. The predominant role of helicases is to unwind DNA or RNA duplexes in advance of the replication machinery. The majority of RNA helicases, termed DExD/H proteins, are not involved in DNA replication but, instead, participate in various aspects of RNA metabolism, including translation initiation, pre-mRNA splicing and ribosome assembly. Helicases act at specific steps in each pathway and are believed to disrupt inter- and/or intramolecular RNA helices and ribonucleoprotein complexes. Helices in natural RNA molecules do not generally exceed 20 bp and the DExD/H helicases are assumed to initiate duplex disruption unidirectionally with a step size of 4–6 bp. Since most RNA duplexes exist as short helices, melting 4–6 bp of a substrate helix may be sufficient to cause the remainder of the helix to spontaneously dissociate. Alternatively, DExD/H proteins disrupt helices by acting directly upon the end or the middle of the helix.

6.4 Regulation at the Level of Translation

During the readout of genetic information into polypeptide chains, translation is the last and probably the least accurate process, though considerable accuracy of protein synthesis is crucial for cell survival. Errors in translation are divided into two types: *missence errors* and *processivity errors*. Missence errors occur when ribosomes accept a noncognate AA-tRNA or an aminoacyl tRNA synthetase mischarges a tRNA with the wrong amino acid. This type of error is limited to a particular amino acid and does not necessarily inactivate the protein product. In contrast, processivity errors include frameshift errors, false recognition of a sense codon by a release factor and drop-off (see Section 6.4.2). This type of mistake often results in truncated products, and in the case of frameshifting, the amino acid sequence incorporated after the shift is wrong. The frequency of missence errors was estimated to be between 10^{-3} and 10^{-4}, while processivity errors were estimated to be in the range of 10^{-4} to 10^{-7}. It has also been noted that processivity errors occur in a sequence-dependent manner and are likely to be more efficient in particular places than in others. Furthermore, it has been discovered that expression of some genes requires nonstandard processivity, called programmed frameshifting (see Section 6.4.5).

Ribosomes

The bacterial 70S ribosome is a large and dynamic ribonucleoprotein machine composed of a large 50S and a small 30S subunit which are assembled at the *t*ranslation *i*nitiation *r*egion (TIR) of mRNAs. Approximately two-thirds of the ribosome consist of RNA and one-third consists of proteins. The 50S large subunit is formed by two RNAs (23S, 5S) and 33 proteins (L1–L36). The 30S subunit is composed of 21 proteins (S1–S21) and an RNA of approximately 1500 nucleotides sedimenting at 16S. It binds mRNA and the anticodon loop and stem of tRNAs. *E. coli* codes for 41 tRNA species with different anticodons. The ribosome must select the tRNA with an anticodon complementary to the codon of the mRNA, the cognate tRNA. The error rate of tRNA selection in the decoding process is 10^{-3} to 10^{-4}. Since there are 41 different tRNAs, but 61 different codons, some tRNAs must be able to decode more than one codon. It could be shown that the codon–anticodon interaction at the first two positions is strictly monitored, whereas the ribosome is able to tolerate noncanonical base pairs at the third position. The 50S large ribosomal subunit is composed of 34 proteins and two RNAs sedimenting at 5S and 23S, respectively. It contains a tunnel about 100 Å long, with an average diameter of 15 Å and starting at the *p*eptidyl*t*ransferase *c*enter (PTC), where formation of the peptide bonds occurs. During the peptidyl transfer reaction, the α-amino group of the A-site tRNA attacks the carbonyl group of the P-site peptidyl group, which is linked to the tRNA via an ester bond. Since the PTC is completely devoid of protein, the ribosome acts as a ribozyme. After the peptidyltransferase reaction has occurred, a deacylated tRNA is left in the P-site and the A-site is covalently bound to the peptide chain. Next, the P-site tRNA moves into the E-site, ready for ejection from the ribosome. Concomitantly, the A-site peptidyl tRNA has to move to the P-site. The ribosome is able to incorporate 10–20 amino acids into the nascent chain per second with an extreme accuracy, making only one misincorporation for every 3000 codons deciphered. This high translational fidelity is attained by carefully monitoring the stereochemistry of the tRNA–mRNA anticodon–codon interaction.

There is overwhelming evidence that the genetic code is, with relatively minor variations, conserved throughout species. It follows that the apparatus that deciphers this code should be correspondingly conserved. Current X-ray diffraction analysis of the eubacterial ribosomes at 5.5-Å resolution has resulted in the direct visualization of the tRNA substrates in the ribosomal A, P and E sites. The ribosomal proteins, the tRNAs and the aminoacyl tRNA synthetases have remarkably conserved structures throughout the species, as have the protein factors that are involved in initiation, elongation and termination.

mRNA

The mRNA interacts specifically with tRNA as well as the 30S ribosomal subunit during translation initiation. The mRNA sequence covered by the ribosome in this phase is called the *r*ibosome *b*inding *s*ite (RBS) and extends over about 50 nucleotides. Most bacterial mRNAs are normally polycistronic and each open read-

ing frame is flanked by signals for the initiation and termination of protein synthesis. Upstream from the initiation codon is the 5′UTR which contains the SD sequence AGGAGGU in *E. coli*, or related sequences thereof, which undergoes basepairing to the 3′ end of 16S rRNA of the 30S subunit. The SD sequence is located 7±2 nucleotides upstream from the initiation codon and is responsible for anchoring of the 30S ribosomal subunit on the mRNA. A direct consequence of the SD interaction is the adjustement of the initiation codon to the ribosomal P-site, where it interacts with fMet-$tRNA_f^{Met}$. Initiation codons can be AUG, GUG or UUG, occurring in *E. coli* at a frequency of 90, 8 or 1%, respectively. The exceptional AUU initiation codon has been observed in *infC* (encoding IF3) and *pcnB* [encoding poly(A) polymerase]. The ribosomal protein S1 interacts with a pyrimidine-rich region 5′ to the SD region on mRNAs. A region downstream from the initiation codon of several mRNAs was found to exhibit complementarity to bases within helix 44 of the 16S RNA. This region was named the *d*ownstream *b*ox (DB) and the complementary region within the 16S RNA anti-DB. But there is no genetic and biochemical evidence in support of the proposed role of the DB versus anti-DB interaction.

Leaderless mRNAs

While most mRNAs are canonical with a SD sequence, leaderless ones also occur, with ~40 identified cases in bacteria. Leaderless mRNAs start at, or a few nucleotides 5′ upstream of, the initiation codon and their binding to the ribosomes involves a mechanism that is somewhat different from the binding of canonical mRNAs. It was suggested that leaderless mRNA is recognized by a 30S-IF2- fMet-$tRNA_f^{Met}$ complex, as shown schematically in Fig. 6.44. Leaderless mRNAs are mainly found in Gram-positive bacteria and start with an AUG.

Translation is conceptually divided into four phases: initiation, elongation, termination and ribosome recycling.

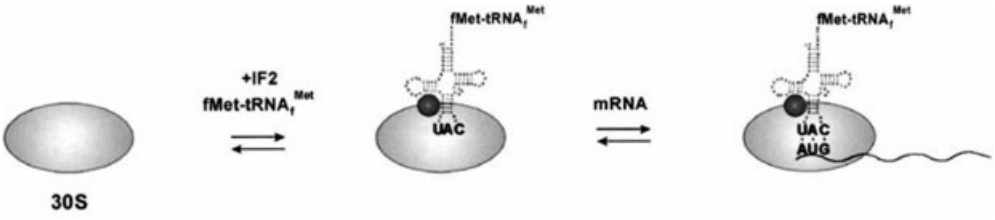


Fig. 6.44 Recognition of leaderless mRNAs by the translation machinery. First, IF2 and the initiator fMet-tRNA form a complex at the 30S ribosomal subunit. Next, the anti-codon of the tRNA will base-pair with the start codon located at the 5′ end of the mRNA. I. Moll, et al. **2002**, *Mol. Microbiol.* 43, 239–246; Fig. 1.

6.4.1 Initiation of Translation

Translation initiation is defined as the sequence of events leading up to formation of the first peptide bond. Components of the initiation complex are the ribosome, an mRNA, the initiator tRNA fMet-$tRNA_f^{Met}$, the three initiation factors, IF1, IF2 and IF3, and one molecule of GTP. A complex of a 30S small ribosomal subunit with all three initiation factors binds mRNA and fMet-$tRNA_f^{Met}$ in random order. Next, this complex must select the start codon defining the beginning of the reading frame on the mRNA. This selection entails decoding of the initiator fMet-$tRNA_f^{Met}$ by the initiation codon and is achieved through a process kinetically controlled by the three initiation factors. The resulting *30S initiation complex* can either dissociate into its individual components or be stabilized by its association with the 50S ribosomal subunit to form the *70S initiation complex*. The initiation factors IF1 and IF3 are ejected from the 30S subunit during 70S initiation complex formation; and this process activates the intrinsic GTPase activity of IF2 and results in the adjustment of fMet-$tRNA_f^{Met}$ at the ribosomal P-site. During this process, IF2 is also ejected and allows the A-site binding of the elongator aminoacyl-tRNA encoded by the second mRNA codon. This aminoacyl-tRNA is targeted by EF-Tu in the ternary complex also containing GTP. The subsequent formation of the first peptide bond between this amino acid and the *N*-formyl-methionine yields the initiation dipeptide. With the subsequent first translocation event, translation enters the elongation phase.

The recognition and binding of the 30S ribosomal subunit to the correct start site of the transcript depends on the structural elements of the TIR of the mRNA. Aside from the leaderless mRNAs (see above), the TIR of canonical mRNAs contains, upstream of the initiation codon, a 5′UTR of variable length and structure. The 5′UTR includes a purine-rich Shine–Dalgarno sequence which is complementary to a sequence at the 3′ terminus of the 16S rRNA. The Shine–Dalgarno sequence and the start codon are separated by a spacer region with an optimum spacing of about eight nucleotides. When the spacing drops below four or above 14, the translation rate decreases by more than ten-fold. The optimal combinations of these TIR elements can maximize translation of mRNA. The start codon is usually AUG or GUG as already mentioned; and, rarely, other codons such as UUG (is used to initiate ribosomal protein S20), AUU and AUA can also initiate synthesis. Start codons are ambigous, since they also specify the insertion of methionine, valine, or other amino acids at internal positions. These triplets also occur out of phase in the genetic transcript. Therefore, both the start codon and the Shine–Dalgarno sequence guarantee the phase translation of the genetic transcript.

What is known about the function of the three IF factors? Formation of the 30S-mRNA-fMet-tRNA complex requires binding of the IF2-GTP-fMet-tRNA into the P-site. IF3 prevents the association of ribosomal subunits and impedes the codon-specific attachment of elongator tRNAs. IF2 is a latent GTPase that may initiate these proofreading events. IF2 may discriminate the initiator tRNAs from the elongator tRNAs by recognizing the acceptor end of the tRNA. In contrast, IF3 ap-

pears to select the anticodon stem-loop of the initiator tRNA. The 71-amino-acid IF1 stimulates the action of IF2 and IF3 by binding to the A site of the 30S subunit. Thus, the initiation factors guarantee the exclusive occupancy of the P-site by fMet-tRNA. The initiation factors are probably not involved in mRNA binding, but instead affect the position of the mRNA on the ribosome. The binding of the 180-amino-acid IF3 results in a conformational change in the ribosomes, which promotes the subsequent binding of the mRNA to the 30S subunit.Two forms of IF2 and IF3 occur in *E. coli* cells. It is possible that these forms act on alternative pathways in which fMet-tRNA interacts first with the 30S particle or with the 30S-mRNA complex.

Does the DNA sequence immediately downstream of the start codon influence the translation efficiency? Here, rare codons play a role. The rare codons AGA/AGG as well as some other minor codons appear to be preferentially used within the first 25 codons in *E. coli*. Consecutive AGG or AGG AGG codons near the 5′ end of the message give significantly lower gene expression, even though these codon sequences further down in the gene had no apparent effect on gene expression. A recent survey of the effects on gene expression by different codons at position +2, immediately following the initiation codon, has shown an effect that varies ~20-fold. This +2 codon influence on gene expression is the result of an evolutionary adaptation, since codons giving high or low expression tend to be over- or under-represented in natural genes, respectively. Thus, the codon composition in the early coding region is one of the determinants for the apparent efficiency of gene expression at the translational level.

6.4.2 Elongation of Translation

Protein chain elongation entails a cycle consisting of aligment of aminoacyl tRNAs by their specific codons in mRNA (decoding), peptide bond synthesis and movement of mRNA relative to the ribosome (translocation). The eubacterial proteins EF-Tu, EF-Ts and EF-G faciliate these processes on ribosomes.

The elongation phase of protein synthesis is promoted by two G proteins, elongation factor EF-Tu, which delivers aminoacyl tRNAs to the ribosome, and EF-G, which catalyzes the translocation step, during which the A- and P-site tRNAs move to the P and E sites of the elongating ribosome, respectively, and mRNA is advanced by one codon. EF-G binds to the ribosome in its GTP form, hydrolyzes GTP to drive tRNA movement on the ribosome and is released in its GDP form. The functional cycle is completed upon GDP release and reactivation of the empty factor by binding of a GTP molecule. The first elongator tRNA is attached to the ribosomal complex as a GTP-EF-Tu-aminoacyl-tRNA intermediate, which accelerates the rate of binding of the aminoacyl tRNA moiety to the mRNA-programmed ribosome. Hydrolysis of GTP, promoted by L7/L12 of the 50S subunit, is required before a peptide bond can be made. The energy of GTP cleavage may be expended to proofread near-cognate aminoacyl tRNAs and/or allow for the alignment of the correct codon–anticodon interaction. After GTP hydrolysis, EF-Tu-GDP leaves the

ribosome and a peptide bond can be formed. EF-Ts catalyzes the exchange of GTP in EF-Tu-GDP with free GTP. Following peptide bond formation with the peptidyl tRNA, the tRNA must be moved or translocated from the ribosomal A- to the P-site. Peptide bond synthesis by the peptidyl transferase (can be inhibited by sparsomycin) of the 50S particle also leaves the deacyl tRNA in the P-site of the ribosome. Thus, after or during translocation, the deacyl tRNA must also be translocated from the P- to the E-site. The movement (translocation) of the nascent peptidyl tRNA from the A- to the P-site is inherent to the ribosome but is catalyzed by elongation factor EF-G and GTP. This movement is substantial, involving changes in the order of 50 Å at the elbow of the tRNA during each step of the elongation cycle. The tRNA substrates must remain bound to the ribosome during this process in order to maintain the reading frame. The energy to drive this mechanism comes largely from hydrolysis of GTP that results from contact of the EF-G-GTP complex with the ribosome. Presumably, GTP hydrolysis allows the release of EF-G from ribosomes.

When ribosomes translate a polycistronic mRNA, each coding region is preceded by a Shine–Dalgarno sequence and flanked by a start and a stop codon. Ribosomes translate these coding region by binding independently to the different Shine–Dalgarno sequences and dissociating at the stop codons. There is one exception to this rule where one ribosome translates to adjacent open reading frames without dissociating after translation of the first one. This situation is designated *translational coupling*. Two genes are said to be translationally coupled if translation of the upstream gene is required for translation of the immediate downstream gene. Translational coupling is a mechanism by which an appropriate quantitative ratio of a certain protein to another protein is obtained and can be classified into three catagories. In the first category, the downstream cistron becomes available for translation by free ribosomes through the action of the ribosomes translating the upstream cistron (hereafter called upstream ribosomes). However, the downstream gene is translated exclusively by free ribosomes. In the second category, the downstream gene becomes available in the same way, but both the free and the upstream ribosomes translate the downstream cistron. In the third category, coupling occurs by a reinitiation mechanism in which ribosomal subunits that have translated the upstream cistron are reused at the start of the next cistron. In these cases, the end of the first cistron often overlaps the start of the second cistron (e.g., AUGA).

Two mechanisms compete with unimpaired elongation: drop-off and SsrA-mediated tagging.

6.4.2.1 **Drop-off Mechanism**

Drop-off is a mechanism which involves spontaneous dissociation of the peptidyl tRNA from the ribosome before hydrolysis of the ester bond. The peptidyl tRNA released by drop-off is then hydrolyzed by the enzyme *p*eptidyl *t*RNA *h*ydrolase encoded by the *pth* gene, allowing the tRNA to be charged by the cognate aminoacyl tRNA synthetase and reutilized in protein synthesis. This premature dissociation

of peptidyl tRNA from the translating ribosome occurs normally during protein synthesis; and the *pth* gene is essential. Two set of proteins catalyze drop-off. One set consists of three factors involved in ribosome recycling: the termination factor RF3, the *r*ibosome *r*ecycling *f*actor RRF and the elongation factor EF-G, where the latter primarily catalyzes the translocation of the peptidyl tRNA from the A- to the P-site. *In vitro*, these three proteins stimulate drop-off of dipeptidyl tRNA at least 30-fold. The second set of proteins involves the initiation factors IF1 and IF2. Mutations have been isolated which reduce the level of drop-off. They reduce the level of expression of RRF or inactivate the gene coding for RF3. When the frequency of drop-off exceeds the capacity of Pth to recycle the tRNA sequestered as peptidyl tRNA, starvation for essential tRNA isoacceptors can occur, leading to an inhibition of protein synthesis and eventually cell death. In accordance with this observation, a complete shutdown in protein synthesis takes place when a temperature-sensitive *pth* mutation is shifted to 43 °C. During translation of *lacZ*, aborted intermediates are ~40% as frequent as completed chains. For a protein of average size, a similar drop-off frequency would result in truncated chains ~10% as often as the full-length protein. Such failures in processivity represent an enormous energetic loss for the cell, but are believed to be an unavoidable consequence of optimizing the level of translation accuracy.

Two interesting observations illustrate the significance of drop-off as a putative regulation mechanism in protein synthesis and both involve the bacteriophage λ, though both examples are somewhat artificial. Multicopy expression of genes that contain the Arg triplets AGA and AGG at positions very close to the initiation codon inhibits cell growth and protein synthesis. An example of this type of inhibition follows the expression of the λ *int* gene, which harbors AGA and AGG triplets at positions 3 and 4. It has been proposed that *int*-mediated inhibition results from ribosome stalling at rare Arg codons and subsequent dissociation of peptidyl $tRNA^{Arg4}$. If a *pth* mutants are infected with phage λ, they are unable to maintain normal bacteriophage vegetative growth, due to the expression of very short, λ-encoded ORFs. These ORFs, also called mini-genes, consist of three codons only: a start, a sense and a stop codon encoded by the *bar* genes. When one of the *bar* genes is carried on a multicopy plasmid, induction of the mini-gene can kill *pth* mutants. Transcription of the *barI* and *barII* loci yields transcripts containing Shine–Dalgarno sequences followed by the short coding region. These transcripts are translated, producing fMet-tRNA-Ile-$tRNA^{Ile}$, but that dipeptide release is partially defective, leading to dipeptidyl $tRNA^{Ile}$ drop-off and starvation for $tRNA^{Ile}$. Four parameters have been described contributing to the growth inhibitory effect. First, replacement of the relatively weak ribosome binding site of the natural *bar* loci by a strong Shine–Dalgarno sequence causes cell death in the presence of the wild-type allele. This is due, at least in part, to the increased probability that the ribosomes retranslate the same mini-mRNA without dissociating from it after dipeptide release. The other factors affecting the rate of peptidyl tRNA accumulation are the termination efficiency, the rate of drop-off and the rate of hydrolysis of the resulting peptidyl tRNA by the Pth hydrolase. They vary according to the sense codon, the stop codon present and the nucleotide following the stop codon.

6.4.2.2 The SsrA-tagging Mechanism (*Trans*-translation)

All organisms have evolved quality control mechanisms to avoid or to deal with errors that occur during protein synthesis. While truncated transcripts lacking stop codons (*nonstop mRNAs*) are rapidly degraded in eukaryotic cells reducing the synthesis of incomplete proteins, bacteria have developed a specialized RNA designated SsrA or tmRNA. In *E. coli*, nonstop mRNAs have been shown to arise by premature termination of transcription, by readthrough of termination codons during translation and by endonucleolytic cleavage upstream of a stop codon. In addition, tmRNA has also been shown to function during translation of full-length messages when ribosomes pause or idle at rare codons or stop codons because the cognate tRNA or protein release factor is scarce.

What happens when translating ribosomes arrive at the 3′ end of a mRNA molecule devoid of a stop codon? Will they fall off or will they become stalled? If they stay bound at the 3′ end, will they stay forever or at least for some time? Stalled ribosomes will be rescued by the tmRNA which acts both as a tRNA and as a mRNA. This tmRNA adopts a partial structure of a tRNA molecule, binds to alanine tRNA synthetase where it is charged with alanine. When ribosomes are stalled and the A-site stays empty for a while, the charged tmRNA will bind into the empty A-site aided by the SmpB protein (Fig. 6.45). Next, the tmRNA is covalently linked to the nascent polypeptide chain through its alanine residue followed by moving to the P-site. Then, the ribosome dissociates from the mRNA and binds to a codon called *resume codon* which is part of an open reading frame lo-

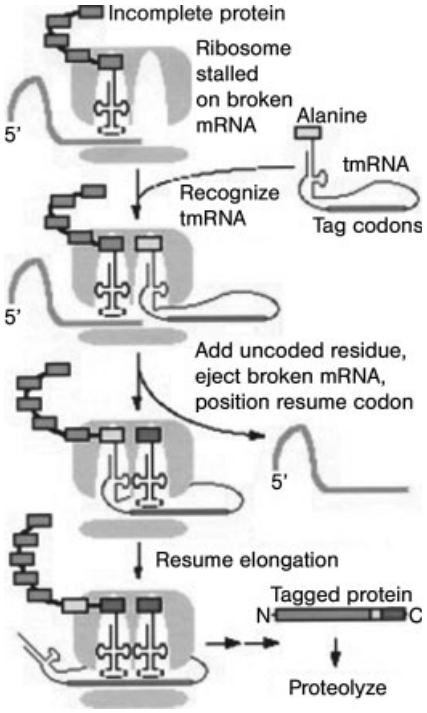


Fig. 6.45 Mechanism of tmRNA-mediated ribosome release and nascent polypeptide tagging. The ribosome is stalled at the 3′ end of of a mRNA devoid of its stop codon. The empty A-site is recognized by the transfer-like part of the tmRNA which binds into this site and transfers the alanine residue to the nascent polypeptide chain. Next, the ribosome jumps to the open reading frame present within the tmRNA, continues translation and is released at the stop codon positioned at the end of the open reading frame. The amino acid residues added to the nacent polypeptide chain act as a proteolytic tag which is normally recognized by the SspB adaptor protein targeting these proteins to the ClpXP protease. A.W. Karzai, et al. **2000**, *Nat. Struct. Biol.* 7, 449; Fig. 1.

cated on the tmRNA. Translation continues on the tmRNA (this process is designated *trans*-translation) until a stop codon is reached where termination of translation occurs. In *E. coli*, a total of 11 codons are translated, resulting in the addition of 11 amino acid residues beyond the alanine. Whenever a stalled ribosome is rescued by the SsrA-tagging mechanism, identical amino acids are added to the polypeptide chain; and the number and composition of these amino acids vary from species to species. These amino acids serve as a proteolytic tag which is recognized by an ATP-specific protease. In *E. coli*, ClpAp, ClpXP and FtsH can degrade proteins with the SsrA-tag, but most of these molecules are escorted to ClpXP by the adaptor molecule SspB (see also under Section 7.4.1). In summary, the tmRNA fullfils two important functions:

- It rescues stalled ribosomes.
- It adds a proteolytic tag to these normally truncated polypeptides allowing their subsequent rapid degradation.

In *C. crescentus*, as in all other α-proteobacteria, the SsrA RNA is composed of two RNA molecules due to a circular permutation in the *ssrA* gene. The *ssrA* gene is transcribed as a single RNA, pre-SsrA, which is predicted to form a similar secondary structure to the mature SsrA except that the tRNA-like 5′ and 3′ ends are connected by a closed loop. This loop is then excised to produce mature SsrA, composed of the coding RNA and the acceptor RNA.

6.4.3 Termination of Translation

Termination of translation, the release of the newly synthesized polypeptide chain, occurs when the ribosome encounters a stop codon on mRNA in its current reading frame. Eubacteria (and eukaryotes) use the three stop codons UAG (amber), UAA (ochre) and UGA (opal). In eubacteria, two class-I peptide *r*elease *f*actors (RF) are necessary to decode these signals: RF1 reads UAG and UAA codons, and RF2 reads UGA and UAA codons. In eukaryotes, a single omnipotent release factor (eRF1) recognizes all three stop codons. Two additional proteins in prokaryotes, class-II factors RF3 and *r*ibosome *r*ecycling *f*actor RRF are involved in later stages (see below). The codon specificity of the two RFs is not well understood at the molecular level. Functional studies have identified one domain as important for peptide release and another for codon recognition. The first contains the tripeptide Gly-Gly-Gln (GGQ) and this motif is perfectly conserved among RFs from all organisms. GGQ interacts with the peptidyl transferase centre on the 50S ribosomal subunit, activating hydrolysis of the ester bond that links the nascent polypeptide to tRNA on the ribosome. The second domain contains a tripeptide motif that determines the identity of bacterial RFs. The tripeptide Pro-Ala-Thr (PAT) in RF1 and Ser-Pro-Phe (SPF) in RF2 determines whether the RF reads UAA and UAG or UAA and UGA; and the first and third residues of the tripeptide have been proposed to interact directly with the stop codon on the ribosome. How the two factors recognize the uridine common to the three stop codons remains elusive.

The relative efficiency of decoding of stop codons depends on their immediate nucleotide context. Experiments point to the –4 nucleotide in an extended stop codon as the most important determinant of termination efficiency. The extended UAAU stop signal is thought to be the most efficient translation termination signal in *E. coli*. This led to the suggestion that RF1/RF2 might in fact recognize an extended tetranucleotide stop signal.

In vivo and *in vitro* observations suggest that UGA can be read as termination, tryptophan, cysteine or selenocysteine without loss of viability. This misreading is necessary for the synthesis of some essential proteins. There is genetic and biochemical evidence that the code to terminate synthesis is longer than a single triplet and that, in certain cases, sequences in the mRNA either 5′ or 3′ of these codons may influence the decision of termination or insertion of an amino acid (suppression). It has been suggested that RF1 and RF2 interact directly with nonsense codons in mRNA because the release factors cross-link to the stop codons in mRNA.

Have all components of the translation machinery been identified? Reconstitution synthesis using all of the homogeneous initiation, elongation and termination factors has revealed that they do not suffice to promote the synthesis of peptides directed by native mRNA templates. Three different proteins are missing in this *in vitro* translation experiment. One of the proteins, EF-P, is required for peptide bond synthesis by 70S ribosomes in the absence of organic solvents. The second protein (gene: *deaD*) is the W2 (IF4A) helix-destabilizing protein, which is strongly induced under cold stress. It stimulates protein synthesis, programmed by templates that harbor secondary structures. Formation of the ternary initiation complex, fMet-tRNA-30S-mRNA, is markedly dependent on the IF4A helicase. It is possible that IF4A protein unwinds structures that occur in the start or in the coding region so as to effect proper accomodation of the initiation complex. Thus, IF4A is an important factor for initiation on structered mRNA sites. A third required protein, now called RbbA (for *rib*osome-*b*ound *A*TPase) accounts for the bulk of the ATPase activity associated with 70S ribosomes and 30S subunits. The primary amino acid sequence of RbbA reveals two ATP/GTP-binding motifs GX4GK(S/T). The exact function of RbbA in protein synthesis has not yet been determined, but several properties have been described: (a) RbbA has a site for EF-Tu-GDP; and (b) RbbA is bound to both 70S ribosomes and 30S subunits. RbbA also binds specifically to 16S rRNA suggesting a role for RbbA in the decoding process of protein synthesis.

L. Kisselev, et al. **2003**, Termination of translation: interplay of mRNA, rRNAs and release factors? *EMBO J.* 22, 175–182.

6.4.4
Ribosome Recycling

Ribosome recycling is the final stage of translation and involves the concerted action of RRF and EF-G to disassemble the post-termination complex for the next round of translation. RRF is universally conserved in bacteria, but is not present

in archaea or eukaryotes (with the exception of chloroplast and mitochondrial RRFs). Deletion of *frr*, the gene encoding RRF, is lethal to *E. coli* cells and, in the absence of RRF, ribosomes remain bound to the mRNA and initiate spontaneous translation downstream of the stop codon. Although RRF was discovered in the early 1970s, the exact mechanism by which RRF mediates ribosome recycling still remains to be fully elucidated. RF3 catalyzes the dissociation of RF1 or RF2 from the ribosomal A-site, leaving a post-termination ribosome complex containing mRNA, deacylated tRNA in the P-site and an empty A-site (Fig. 6.46). This post-termination complex is then disassembled by the action of RRF in concert with

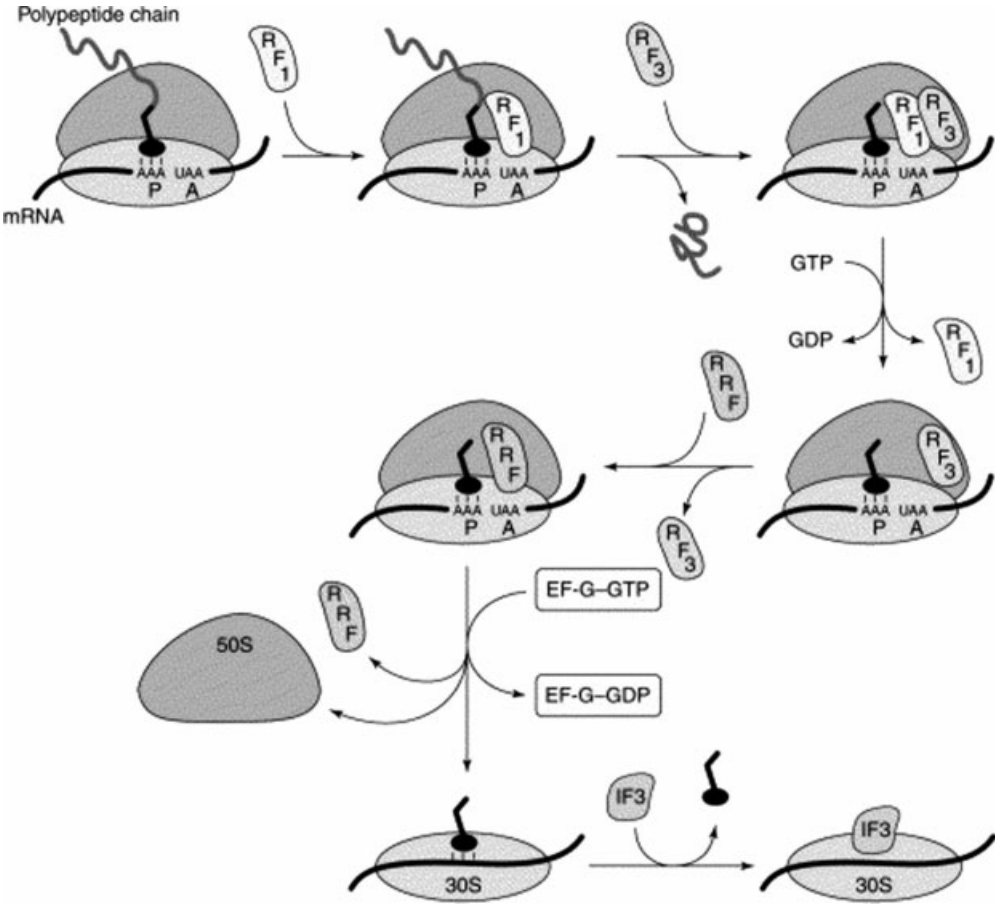


Fig. 6.46 Termination of translation and ribosome recycling in *E. coli*. Termination of translation starts with binding of the appropriate release factor (here RF1) to the stop codon in the empty A-site. Next, RF3 will interact with RF1 and displace this factor upon hydrolysis of GTP to GDP. In the third step, RRF will replace RF3 and after interaction with EF-G and GTP hydrolysis, the ribosomes will be dissociated into their subunits. L.L. Kisselev, R.H. Buckingham **2000**, *Trends Biochem. Sci.* 25, 561–566; Fig. 3.

elongation factor EF-G. RRF binds to the A/P-site of the ribosome, but an additional site for RRF ribosomal binding has been reported. It has been proposed that RRF is moved from the A/P-site to this second site by EF-G, resulting in the release of tRNA from the ribosome. This recycling step is important to avoid interference between successive ribosomes translating an mRNA and to maintain a pool of ribosomes available for translation initiation. Two contrary arguments have been made for the disassembly reaction: one argues that the post-termination ribosome is disassembled into a free 70S ribosome, a deacylated tRNA and the mRNA transcript; and the other argues the generation of a free 50S subunit, with the 30S subunit remaining complexed with the mRNA and the deacylated tRNA, followed by complete disassembly by initiation factor IF3. Many details of the sequential events responsible for the disassembly of the post-termination complex remain to be established.

L.L. Kisselev, R.H. Buckingham **2000**, Translation termination comes of age, *Trends Biochem. Sci.* 25, 561–566.

L. Kisselev, et al. **2003**, Termination of translation: interplay of mRNA, rRNAs and release factors? *EMBO J.* 22, 175–182.

6.4.5
Translational Control

Gene regulation was first shown to occur at the level of transcription, later at the level of translation. The main target for translational control is the initiation of protein synthesis, which is generally the rate-limiting step. Regulation can occur at any step of translation initiation, e.g., the formation of the 30S-mRNA binary complex or the conformational change required to promote the codon-anticodon interaction, leading to the active initiation of the 30S-mRNA-tRNA complex. Regulation of translation can occur via activation or repression and involve a *trans*-acting factor (RNA or protein) or a *cis*-acting RNA sequence.

Translation Activation Mechanisms

Three different principles have been described all preventing formation of a secondary structure near the 5′ end of the mRNA sequestering the Shine–Dalgarno sequence which becomes unavailable for binding the ribosomal 30S subunit: activator RNA, activator protein and *cis*-acting RNA (Fig. 6.47). Activator RNAs belonging to the group of small ncRNA (see Section 6.3) bind to the mRNA upstream of the Shine–Dalgarno sequence and prevent formation of a stem-loop structure sequestering the Shine–Dalgarno sequence. One example is binding of the DsrA small ncRNA to the *rpoS* mRNA. While the 5′ end of the *rpoS* transcript folds back on its own RNA and forms a hairpin structure including the Shine–Dalgarno sequence, interaction with DsrA to the 5′ end distracts the inhibitory part from RBS.

Activator proteins prevent formation of the hairpin structure comparable to the activator RNAs. This is exemplified by the 7.4-kDa Com (for *c*ontrol of *mom*) pro-

tein encoded by the phage Mu. The *com* gene forms a bicistronic operon with *mom* (*m*odification *o*f *M*u). *mom* codes for an enzyme which modifies the viral genome, protecting it against a variety of restriction endonucleases. The 62-amino-acid Com protein contains a zinc finger-like structure involved in coordinating four cysteine residues and binds the *com*/*mom* transcript 5′ to the *mom* open reading frame, whose translation start signals are contained in a stem-loop translation-inhibition structure. Com binding to its target site results in a stable change in RNA secondary structure that exposes the translation start signal as shows schematically in Fig. 6.47.

Furthermore, formation of the hairpin can be temperature-dependent. At low temperatures (around 30 °C), the hairpin forms and impairs binding of the 30S ribosomal subunit. After infection of a mammal (37 °C; *prfA* transcript encoded by *L monocytogenes*) or a heat shock (42 °C; *rpoH* transcript encoded by *E. coli*), the increased temperature destabilizes the hairpin and allows translation (see also under Section 9.2).

Some genes are preceded by sequences which have been termed *translational enhancers*. The most prominent example is the enhancer identified in front of the *atpE* gene of *E. coli*. This gene belongs to an operon coding for the proton-translocating ATP synthase composed of eight different types of subunits. The enzyme consists of two parts, F_1 and F_0. The F_1 part of the enzyme carrying the ATP hydrolase activity is composed of five different subunits, while the membrane-embedded F_0 part comprises three types of subunits. While most subunits are present in 1–3 copies, subunit c encoded by *atpE* is present in 10–15 copies. All genes of the *atp* operon are transcribed into a single polycistronic transcript and high-

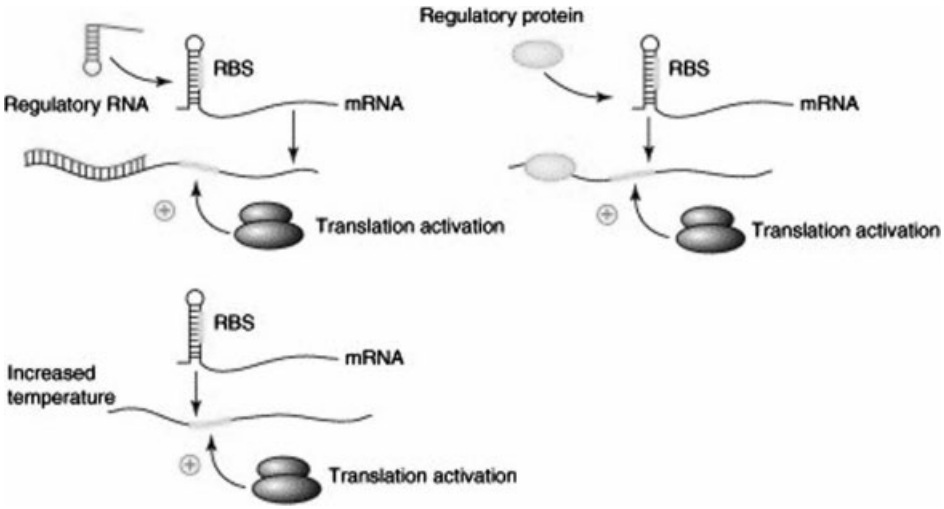


Fig. 6.47 Regulation of translation initiation by activation. A regulatory RNA (left) or protein (middle) binds upstream of the Shine–Dalgarno sequence preventing formation of a secondary structure sequestering the RBS. Formation of the secondary structure is temperature-dependent (right). P. Romby, M. Springer **2003**, *Trends Biochem. Sci.* 19, 155–161; Fig. 1.

level production of the c subunit is obtained post-transcriptionally. It could be shown that an intergenic sequence stretching >20 nucleotides upstream of the Shine–Dalgarno sequence is responsible for high-level translation of *atpE*. It remains an open question how this enhancer sequence stimulates translation of just the downstream open reading frame. Fusing this enhancer sequence to the coding sequence for human interleukin 2 and β-interferon increased translation of both eukaryotic genes 6- to 10-fold.

Translational Repression Mechanism

Repression of translation can be exerted by two different *trans*-acting molecules: anti-sense RNAs and repressor proteins, both interfering with initiation of translation (Fig. 6.48). The anti-sense RNA can be complementary to the mRNA along its whole length (rare) or only partly complementary. Two of the numerous cases where an anti-sense RNA inhibits translation are *micF* and DsrA. The *E. coli* 93-nucleotide *micF* RNA, the first identified chromosomally encoded anti-sense RNA, is induced by stress conditions, including elevated temperature, high osmolarity and redox stress (see also under 6.2.8.5 and 9.3.1). The target of *micF* is the *ompF*-RNA coding for the outer membrane porin OmpF. *micF*-*ompF*-RNA interaction leads to the formation of a partial duplex, which inhibits *ompF* translation and promotes *ompF*-mRNA degradation. The second example involves inhibition of translation of the *hns* mRNA (coding for the histone-like protein H-NS) by DsrA antisense RNA. The 87-nucleotide DsrA-RNA acts at different targets. It inhibits *hns*-mRNA translation by destabilization of the mRNA and by blocking its ribosome binding site.

Translational repressor proteins can act by one of two mechanisms called the *direct competition model* and *entrapment model*. In the direct competition model, the repressor protein and the ribosome compete for the Shine–Dalgarno sequence. Examples are the bacteriophage MS2/R17 coat proteins, which repress translation of the replicative gene, and ribosomal proteins such as S7, S8, L20 and L1, which interact with their own mRNA and inhibit translation of their genes and the addi-

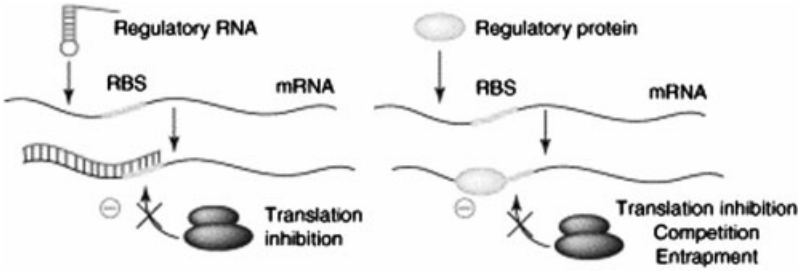


Fig. 6.48 Regulation of translation initiation by repression. Interaction of ribosomes with the RBS can be prevented either by an antisense RNA (left) or by a regulatory protein (right). P. Romby, M. Springer **2003**, *Trends Biochem. Sci.* 19, 155–161; Fig. 1.

tional genes in their operon (see also under 6.3.1). In the entrapment model, the sites on the mRNA recognized by the regulatory protein and the ribosome are distinct; and the regulatory protein traps the 30S subunit in a dead-end complex. One example is the ribosomal protein S4 which represses synthesis of the four ribosomal proteins (including itself) in the *E. coli* α operon. Sequences near the 5′ end of the mRNA fold into a nested pseudoknot structure surrounding the ribosome binding site for the first gene of the operon. This pseudoknot exists in an equilibrium between two conformers with different electrophoretic mobilities. S4 selectively binds to one form of the transcript, while binding of the 30S subunit is nearly equal in the two forms. When S4 and the 30S subunit form a ternary complex with the mRNA, this dead-end complex is unable to bind $tRNA_f^{Met}$. This repression strategy has two novel aspects. First, S4 functions as an allosteric effector to shift an equilibrium between two RNA conformational states. Second, S4 acts at a second, irreversible initiation step to stabilize an inactive 30S subunit-mRNA complex, a mechanism called entrapment.

P. Romby, M. Springer **2003**, Bacterial translational control at atomic resolution, *Trends Genet.* 19, 155–161.

P.J. Schlax, D.J. Worhunsky **2003**, Translational repression mechanisms in prokaryotes, *Mol. Microbiol.* 48, 1157–1169.

6.4.6

The 21. and 22. Amino Acid Selenocysteine and Pyrrolysine

Selenocysteine

The amino acid selenocysteine (Sec) has been identified as a component of selected proteins in eubacteria, archaea and eukarya. It is mainly found in the active site of oxidoreductases. where it is directly involved in catalysis. The pK_a for Sec is much lower than for Cys (5.2 vs 8.3). Consequently, at physiological pH, the selenol of Sec is mainly in its anionic selenolate form, while the thiol of a cysteine residue is typically protonated, making Sec more reactive than cysteine. Sec is incorporated into polypeptide chains during translation elongation and is encoded by an in-frame UGA codon interrupting the reading frame of the respective gene followed by a hairpin structure, the SECIS (*se*leno*c*ysteine *in*sertion) element. Both Sec determinants have to be recognized for reading through the stop codon instead of terminating translation. The UGA codon is recognized by $tRNA^{Sec}$, whose UCA anticodon is complementary to the stop codon. There is a specific translation elongation factor, SelB, binding Sec-$tRNA^{Sec}$ in a GTP-dependent manner and delivering it to the ribosomal A-site only in the presence of the *cis*-acting SECIS element. SelB binds directly to the SECIS element via a C-terminal extension called domain IV, which is absent in EF-Tu. In *E. coli*, there are three genes coding for selenium-containing formate dehydrogenases expressed either constitutively or under anaerobic conditions. The biosynthesis of selenocysteine and its insertion into these proteins require the function of at least four enzymes. The *selC* gene encodes the $tRNA^{Sec}$, which is charged with L-serine by seryl-tRNA synthetase (*serC*) to form Ser-$tRNA^{Sec}$, and serves as the adapter at which the seryl

moiety is converted into the selenocysteyl derivative. This conversion reaction is catalyzed by selenocysteine synthase (*selA*) with the aid of monoselenophosphate as the selenium donor substrate, which is synthesized from selenide and ATP by selenonphosphate synthetase (*selD*). The *selB* gene codes for a specialized translation factor taking over the function of the elongation factor EF-Tu in selenocysteine insertion. The SelB protein binds two RNA ligands, namely Sec-tRNA^{Sec} and the SECIS (*se*leno*c*ysteine *in*sertion) element of the mRNAs coding for selenoproteins. In eubacteria, this mRNA structure is located immediately 3′ of the UGA codon that determines the position of selenocysteine incorporation in the nascent polypeptide chain. The binding order is random, but interaction with both RNA substrates leads to a stabilization of the resulting quaternary complex consisting of SelB, GTP, Sec-tRNA^{Sec} and the SECIS element.

Pyrrolysine

The 22. amino acid is pyrrolysine (Pyl), another noncanonical amino acid that is found in certain enzymes of Archaea, including methylamine methyltransferases and transposases, where it is incorporated at a UAG codon. The typical tRNA charging reaction occurs in two steps: First, activation of the amino acid with ATP to form an aminoacyl adenylate, and then, reaction of the adenylate with a pyrrolysine-specific tRNA (tRNA^{Pyl}) with an anticodon complementary to UAG, to covalently link the amino acid to the RNA molecule. Both reactions are catalyzed by a specific aminoacyl tRNA synthetase. Pyrrolysine-specific UAG codons in methyltransferase RNAs could be identified by RNA hairpin loops designated PYLIS similar to the selenocysteine insertion element.

6.4.7 Mechanisms to Create Multiple Gene Products From a Single Cistron

In general, the ''one gene, one enzyme'' hypothesis stated by Beadle and Tatum in 1942 reflects the situation in bacteria. But there are rare exceptions where one gene codes for two or even more proteins. A few different cases have been described: (a) in-phase overlapping genes with the use of internal, in-phase translation start sites or readthrough of a termination codon (the latter process also called redefinition of a stop codon), (b) programmed translational frameshifting, (c) bypassing genes, (d) genes within genes, (e) polyprotein processing and (f) moonlighting proteins.

6.4.7.1 In-phase Overlapping Genes with the Use of Internal, In-phase Translation Start Sites or Readthrough of a Termination Codon

Several in-phase overlapping genes have been described which code for two proteins due to an additional internal start codon: (a) the *E. coli* proteins ClpB and CheA, (b) the Tn*5*-encoded transposase and inhibitor and (c) the *B. subtilis* spore coat protein SafA.

Due to the presence of an internal start codon, two versions of the *E. coli* ClpB exist *in vivo*: a full-length protein (aa 1–857; 93 kDa) and an N-terminal truncated variant (aa 149–857; 79 kDa). Six ClpB monomers form a ring-like structure (see Section 7.2.3) and both versions have been demonstrated to form mixed oligomers. Both versions exhibit ATPase activity, but only the 93-kDa form can be activated by casein. The N-terminal portion of the larger ClpB has been suggested to contain one or more sites or domains responsible for protein binding. Mixing of the larger monomer with differing amounts of the shorter version leads to a gradually reduced protein-activated ATPase activity. The CheA protein, which is part of the chemotaxis signal transduction system, also exists in two forms as a consequence of in-frame initiation sites within the *cheA* gene. The full-length form (CheAL) is essential for signal transduction, but the function of the short form (CheAS) is unclear. Tn*5* is a 5.8-kb composite transposon in which two inverted 1.5-kb IS*50* (IS*50*R, IS*50*L) insertion sequences flank three antibiotic resistance determinants. IS50 encodes two proteins, the 476-amino acid transposase (Tnp) and a 421-amino acid *trans*-active inhibitor (Inh) of transposition. Tnp and Inh are read in the same reading frame but differ because Inh lacks the first 55 aa of Tnp. Synthesis of the two proteins is programmed by two different overlapping and possibly competitive promoters. Both proteins can readily dimerize to form inactive heterodimers. Thus, Inh plays a significant role in downregulation of Tn*5* transposition. Tnp and Inh are translated from two different mRNAs, which are affected in opposite ways by the Dam methylase. The upstream Tnp promoter contains two Dam methylation sites overlapping the –10 element making this promoter more active when the replication fork has passed it leaving the –10 region hemimethylated for a few minutes. During endospore formation in *B. subtilis*, over 24 polypeptides are localized to the developing spore and coordinately assembled into a thick multilayered structure called the spore coat. One of these spore coat proteins is the 45-kDa SafA protein encoded by *safA*. In addition, a 30-kDa protein composed of the C-terminal region of SfaA is also found in spore coats whose N-terminal residue is a methionine encoded by codon 164 of the *safA* gene. The 30-kDa form of SafA is produced primarily by initiation of translation at codon 164 of full-length *safA* mRNA transcript. The full-length form of SafA is necessary for formation of an intact coat. The role of the 30-kDa short form is less clear. Blocking accumulation has no detectable effect on spore coat function or morphology, while overexpression blocks sporulation at stage IV.

There is very limited knowledge about readthrough in bacteria. One example is readthrough of the UGA stop codon of the coat protein gene of the single-stranded RNA bacteriophage Qβ to synthesize a much longer protein also required for infectivity. Another example are proteins that are important for biosynthesis and assembly of CS3 pili of CFA/II in enterotoxigenic *E. coli*. These proteins are encoded by the same gene and produced via readthrough of a UAG stop codon.

6.4.7.2 **Programmed Ribosomal Frameshifting**

Natural frameshift errors occur very rarely (it has been calculated to be not higher than one event per 30 000 amino acids incorporated), but programmed ribosomal frameshifting signals increase the probability of tRNA slippage significantly and is used for regulatory purposes (e.g., RF2, some IS elements) or to synthesize an additional useful product in a fixed ratio to that of standard decoding product (e.g., *dnaX*). While in some cases, only low levels (1–2%) of specific frameshifting are important, occasionally up to 50% of the ribosomes change the reading frame. Ribosomal frameshifting occurs at shift-prone sequences; and the proportion of ribosomes that participate in frameshifting is greatly elevated in many cases by recoding signals embedded in the mRNA.

The *prfB* gene of *E. coli* codes for the release factor 2 (RF2) which recognizes and promotes translation termination at UAA and UGA stop codons. The *prfB* gene is interrupted by a UGA stop codon 26 codons downstream of the initiation codon (Fig. 6.49). Frameshifting is dictated by the number of RF2 molecules present within the cell. If the concentration is high, they will efficently terminate translation at the UGA. If its concentration is low, the A-site with the UGA stop codon is left unoccupied for some time allowing the translating ribosomes to carry out a +1 frameshift at the slippery sequence CUUUGAC to synthesize the full-length RF2 protein. The frameshifting process is enhanced in three ways: (a) the relatively poor termination context of the stop codon used here (UGAC), (b) the Shine–Dalgarno sequence located upstream of the stop codon and (c) the similar codon:anticodon contacts formed after the frameshift. Under normal growth conditions, 30–50% of the elongating ribosomes carry out the +1 frameshift.

Frameshifting can be used to produce two proteins of different lengths that have the same N-terminal region and are endowed with different activities. The IS element IS*911* encodes ORFA and ORFB proteins, which, produced without and with –1 translational frameshifting respectively, do not have the same function. While the upstream frame appears to carry a DNA recognition domain, the down-

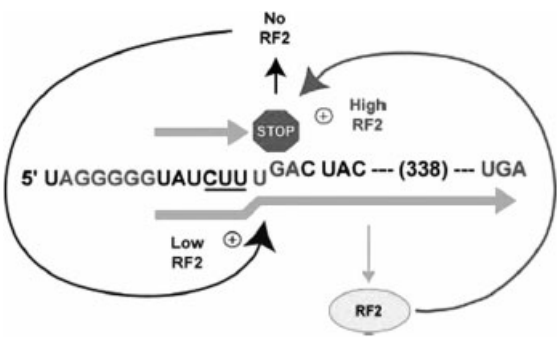


Fig. 6.49 RF2 autoregulates its own synthesis by a +1 frameshift. If the concentration of RF2 is high in the cell, they will terminate translation at the UGA stop codon to result in a 26-amino acid peptide. If the concentration of RF2 is low, elongating ribosomes can carry out a +1 frameshift at the slippery CUUUGAC sequence leading to the full-length RF2 factor. O. Namy, et al. **2004**, *Mol. Cell* 13, 157–168; Fig. 1.

stream frame encodes the catalytic site. The product of the upstream frame alone acts as a modulator of activity, presumably by binding the inverted repeat sequences. A frameshifting event assembles both domains into a single protein, the transposase. The frameshifting rate which involves a –1 frameshift could be influenced by host physiology, thereby coupling transposition activity to the state of the host cell.

In *E. coli*, *dnaX* encodes the τ and γ subunits of DNA polymerase III. While τ is the full-length product encoded by the *dnaX* gene, γ results from a frameshifting event that occurs at two-thirds of the mRNA that directs ribosomes to a premature stop codon (Fig. 6.50). This –1 frameshift occurs on the slippery sequence A AAA AAG by simultaneous slippage of both P and A site $tRNA^{Lys}$ species. This process requires two signals in the mRNA, an SD-like sequence upstream of the slippery sequence and a stem loop structure downstream of it (Fig. 6.50). Both proteins are produced at about equal amounts.

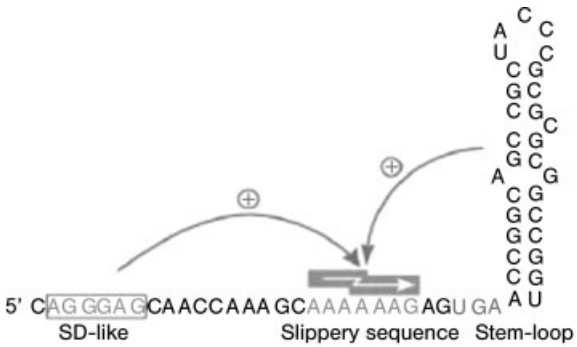


Fig. 6.50 Frameshifting in the *dnaX* gene of *E. coli*. This gene codes for the full length protein τ and the shorter product γ which requires a –1 frameshift at the slippery sequence into the UGA stop codon. O. Namy, et al. **2004**, *Mol. Cell* 13, 157–168; Fig. 4B.

6.4.7.3 **Bypassing**

Bacteriophage T4 gene *60* is an extreme example of bypassing, where 50 noncoding nucleotides (called *coding gap*) separate the two ORFs and must by bypassed in order to synthesize the full-length protein, a topoisomerase subunit. Efficient bypassing of the gene *60* coding gap in *E. coli* requires two matching GGA codons flanking the coding gap which serve as the peptidyl tRNA detachment and repairing sites (also termed as take-off and landing sites). In addition, a UAG stop codon defining the 3′ end of the first ORF, a stem-loop structure containing the stop codon and the take-off site GGA codon within its stem and a *cis*-acting signal in the nascent peptide consisting of a stretch of charged and hydrophobic amino acids encoded by the codons preceding the gap. The suggested mechanism of bypassing has been devided in three steps. In the first step (take-off), the peptidyl tRNA complex dissociates followed by an initial movement of the mRNA prevent-

ing peptidyl $tRNA^{Gly}$ re-pairing with the take-off site GGA. During the second step (scanning), peptidyl $tRNA^{Gly}$ probes the mRNA for a landing site as the mRNA continues to slide through the deoding center of the ribosome. In the third step (landing), peptidyl $tRNA^{Gly}$ pairs with the landing site GGA, resetting the reading frame. In this model, the two GGA codons act as road signs, specifying where bypassing starts and ends. Nearly half of all ribosomes translating gene *60* bypass the coding gap sucessfully while the remaining ribosomes take-off, but their further fate is still unknown.

P.V. Baranov, et al. **2002**, Recoding: translational bifurcation in gene expression, *Gene* 286, 187–201.

O. Namy, et al. **2004**, Reprogrammed genetic decoding in cellular gene expression, *Mol. Cell* 13, 157–168.

6.4.7.4 Genes Within Genes

Gene overlap is common in viruses. One example has been described with bacteriophage MS2. Here, the lysis protein overlaps in a different reading frame with the distal portion of the coat cistron and with the proximal portion of the replicase. Now, such a gene within a gene has been described for bacteria for the first time.

Generally, the gene encoding the protein subunit of RNase P (*rnpA*) is located immediately downstream of the gene coding for the ribosomal protein L34 (*rpmH*). In contrast, both genes overlap in the genus *Thermus*; these genes begin with start codons separated by only 4 bp. While the upstream start codon initiates *rnpA*, the second one initiates the *rpmH* coding sequence, but in the –1 register. The *rnpA* open reading frame continues entirely through *rpmH*. Since L34 is an abundant protein in the cell, whereas the RNase P protein is not, how is the differential expression of both overlapping genes accomplished in *Thermus*? Regulation seems to occur by three different mechanisms:

- The shorter *rpmH* gene is followed by one-to-four potential rho-dependent transcription-termination sites. Translating ribosomes immediately following the RNA polymerase could trigger transcriptional termination at these sites, resulting in a mRNA encoding L34 but not RNase P.
- The *rnpA* gene uses unusual codons primarily found in the region of overlap, reducing the translation rate for *rnpA* by ribosomal stalling.
- The distance of the single ribosome-binding site shared by these two genes is sub-optimal close (3 bp) to the *rnpA* start codon, but optially spaced (7 bp) from the *rpmH* start codon.

J.C. Ellis, J.W. Brown **2003**, Genes within genes within bacteria, *Trends Biochem. Sci.* 28, 521–523.

6.4.7.5 Polyprotein Processing

The proteolytic release of functional proteins from polyprotein precursors is a hallmark of eukaryotes. But the phenomenon of polyprotein synthesis and cleavage also exists in different prokaryotes. Five different examples will be presented.

The first example of a polyprotein concerns the *E. coli* penicillin G acylase which catalyzes the conversion of penicillin G to phenylacetic acid and 6-aminopenicillin acid, a precursor for semisynthetic penicillins. The active enzyme is composed of the two subunits α (23 kDa) and β (69 kDa) and functions as a heterodimer. Both subunits are encoded by the *pac* gene and its open reading frame can be translated into a polypeptide of 846 amino acids. The overall organization of the polyprotein precursor is depicted in Fig. 6.51A. The first 26 amino acids serve as a signal peptide involved in the translocation of the 98-kDa precursor through the inner membrane into the periplasm. This signal peptide is proteolytically removed by the signal peptidase. The next cleavage occurs at the N terminus of the β subunit releasing the free 69-kDa β subunit and the α subunit with an C-terminal extension. Finally, the 54-amino-acid spacer peptide is released. The function of the spacer is to help the folding of the precursor. Whether maturation of the precursor occurs by one or more periplasmic proteases or by intrinsic proteolytic activity remains an open question.

The second example deals with the cytochrome bc_1 complex of *B. japonicum* encoded by *fbcH*. This gene codes for a 687-amino-acid precursor (Fig. 6.51A) which is processed into the 401-amino-acid cytochrome *b* and the 253-amino-acid cytochrome c_1. Both subunits are separated by a 33-amino-acid spacer which shows all the characteristics of a typical signal sequence, except that its location is protein-internal. Details of the maturation reactions are unknown.

Processing of a polyprotein spore coat protein of *B. subtilis* is mentioned as the third example. Endospores of *B. subtilis* are encased in a tough protein shell known as the spore coat. This coat is composed of dozens of proteins, which are organized into an electron-dense outer layer and a thinner, lamella-like inner layer. These layers provide a protective barrier against bactericidal enzymes and chemicals including lysozyme and organic solvents. The *cotF* gene encodes an 160-amino-acid precursor with two cleavage sites (Fig. 6.51A), where both cleavage reactions occur after a GluArg sequence. The mechanism and order of cleavage remains elusive.

The amylase gene of *B. polymyxa* encodes a polyprotein of 1161 amino acids and contains inphase β- and α-amylase coding sequences. The polyprotein starts with a typical signal sequence directing the protein into the medium (Fig. 6.51A). Both amylases are separated by a stretch of about 200 amino acids spacer arranged as a direct repeat of two 104-amino-acid segments. Three β-amylases are produced which have the same N-terminus, but different C-termini (70, 56, 42 kDa). The neutral protease (Npr), the intracellular serine protease (Isp) of *B. polymyxa* and trypsin, chymotrypsin and subtilis are able to process the poylprotein into the respective β-amylases and the α-amylase.

The fifth example is provided by *P. gingivalis*, a Gram-negative, anaerobic bacterium which has been associated with the development of adult periodontitis. Cells

produce three cysteine proteinases, namly RgpA, RgpB and Kgp, that are associated with the cell surface and/or secreted depending on the bacterial strain and environmental conditions. They are involved in pathogenesis by indirectly causing peridontal tissue destruction through the activation of host metalloproteases and by degrading kep proteins and peptides of the immune system. Two genes, *rgpA* and *kgp*, code for polyproteins. Both polyproteins are proteolytically processed to yield an N-terminal proteinase domain and several C-terminal adhesins (Fig. 6.51B).

What selective pressure forced the evolution and maintenance of polyproteins? At least, polyprotein processing guarantees the strict stoichiometric synthesis of polyproteins. Furthermore, within polyproteins, the different domains may fold interdependently, rendering folding independent of chaperones. Third, if the processed proteins form oligomeric structures, diffusion will be a minor problem.

L. Thöny-Meyer, et al. **1992**, Prokaryotic polyprotein precursors, *FEBS Lett.* 307, 62–65.

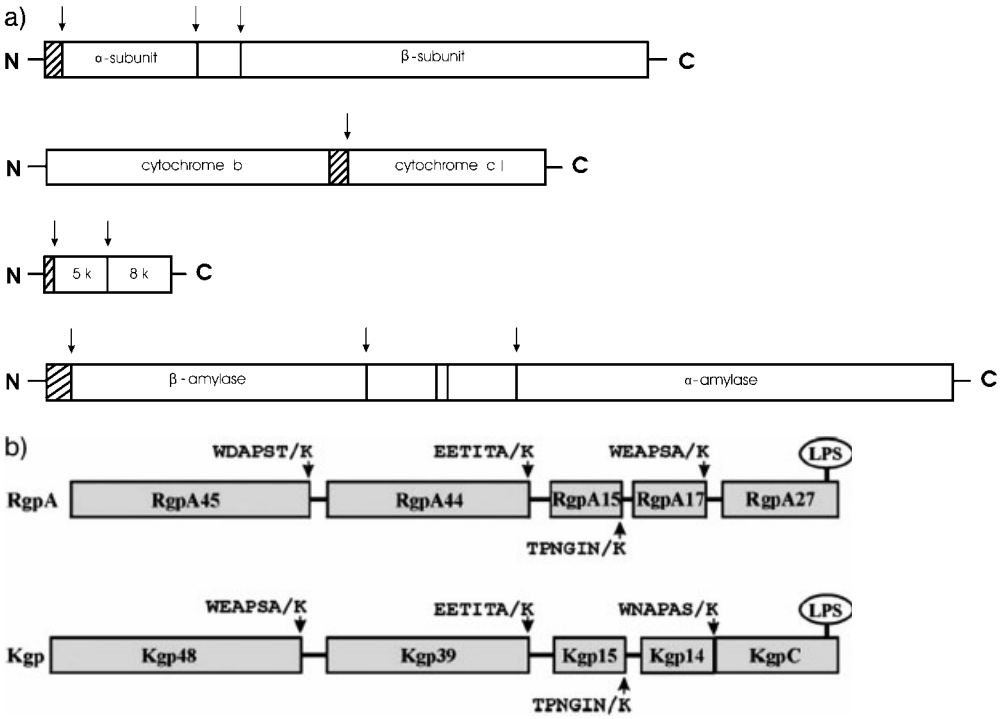


Fig. 6.51 Processing of polyproteins. (A) The entire polyproteins are shown as predicted from their coding region together with the identified processing sites (indicated by vertical arrows): *E. coli* penicillin G acylase, *B. japonicum* cytochrome bc_1, *B. subtilis* spore coat proteins and *B. polymyxa* amylases. (B) *P. gingivalis* RgpA and Kgp polyprotein. Processing yields an N-terminal proteinase domain and four C-terminal adhesins where processing occurs at an X-Lys peptide bond. P.D. Veith, et al. **2004**, *Infect. Immun.* 72, 3655–3657; Fig. 1.

6.4.7.6 Moonlighting Proteins

Moonlighting proteins have more than one function in an organism. There are now numerous examples for moonlighting proteins in all three kingdoms. In *E. coli*, PepA serves as an endopeptidase, but acts also as a DNA-binding protein. BirA adds biotin residues to appropriate proteins and acts as a repressor for the biotin operon. A third example is thioredoxin which reduces ribonucleoside diphosphates to deoxyribonucleoside diphosphates. In addition, the thioredoxin is recruited by T7 phage DNA polymerase. Two general questions can be asked about moonlighting proteins: How did these proteins arise during evolution? How does the cell benefit from these proteins? To answer the first question, we know that many enzymes appear to be much larger than is necessary for performing their known function. They often have large, apparently unused solvent-exposed areas including pockets that can be modified to aquire additional binding sites. If these new binding sites do not interfere with the original one, the cell will benefit from the new function. Alternatively, the different functions of moonlighting proteins might have been spliced together during evolution at the level of DNA or RNA.

C.J. Jeffery **2003**, Moonlighting proteins: old proteins learning new tricks, *Trends Biochem.* 19, 415–417.

6.5 Post-translational Regulation

Post-translational regulation can occur at several levels: (a) proteolysis, (b) polyprotein processing, (c) protein splicing, (d) protein modification and (e) protein cyclization.

6.5.1 Proteolysis

Proteolysis plays an important role in all cellular processes that need to be tightly controlled. This post-translational control is mediated by ATP-dependent proteases (see Section 7.5) that are present and conserved in all organisms.

6.5.2 Protein Splicing

Inteins are genetic elements present within protein-coding sequences with dual functins: protein-splicing and homing endonuclease activities.

6.5.3 Protein Glycosylation

In recent years, experimental evidence has accumulated that protein glycosylation occurs also in eubacteria. In prokaryotes, the earliest examples of protein glycosy-

lation have been found with archaea, which are able to synthesize glycosylated surface layer (S-layer; see Section 1.6) proteins. The S-layer protein of the halophile *H. salinarium* identified in 1976 was the first example of a prokaryotic glycoprotein. Then, S-layer glycoproteins were also identified in *Clostridia*. Later, non-S-layer glycoproteins have been found increasingly in insect and important mammalian bacterial pathogens. Carbohydrates are always attached to the peptide backbone via the amide nitrogen of an Asn residue (*N*-glycosylation) or via the functional hydroxy group of serine or threonine (*O*-glycosylation). The glycoproteins described in pathogenic bacteria appear to be modified primarily by short-chain carbohydrate moieties (mono- to trisaccharides).

Glycosylated S-layer Proteins

A remarkable characteristic of many archael and some eubacterial S-layer proteins is their glycosylation. The glycan chains and linkages are significantly different from those of eukaryotes. For instance, the glycan of *Halobacterium* proteins consists of short, predominantly *N*-glycosidically linked sulfonated heterosaccharides. Bacillaceae S-layers contain identical, repeating units with up to 150 monosaccharide residues, attached primarily by *O*-glycosidic linkages such as β-glucose → tyrosine, β-galactose → tyrosine, or β-*N*-galactosamine → threonine/serine.

Examples for Glycoproteins in Pathogenic Gram-negative Bacteria

In *B. burgdorferi*, the causative agent of Lyme disease, the outer surface proteins OspA and OspB have been identified as the major *N*-glycosylated proteins. In a *Campylobacter* subspecies which causes food-borne illness and diarrhea worldwide, flagellins of a number of strains have been shown to be extensively glycosylated where the carbohydrate moieties are concentrated in a narrow hydrophobic region of the central core domain. Another example is the obligate intracellular pathogen *Chlamydia* subsp. where the 40-kDa major outer membrane protein occurs as a glycoprotein. In the classic human enterotoxic *E. coli* strain H10407, the two autotransporter proteins (see Section 8.6.5) TibA and AIDA-I require glycosylation for activity. Glycosylated proteins have also been described with *Neisseria* and *Mycobacterium* sp. and *P. aeruginosa*.

Glycosylation Machinery

Diverse mechanisms for protein glycosylation have been identified in bacteria. While glycosylation of some proteins requires a glycosylation machinery, others are modified by specific protein glycosyl transferases. One example is the an adhesin heptosyltransferase which adds approximately 19 heptosyl residues to the AIDI-I protein. It is interesting to note that the genes coding for the enzymes involved in glycosylation are those coding for glycosylated proteins are located adjacently, sometimes in glycosylation islands or in the same operon. This clustering is an important prerequisite in that these genes can be transferred together by

horizontal transfer to other strains or even species. Glycosylation has been reported to take place in the cytoplasm, at the inner membrane or in the periplasm via a lipid-linked intermediate, depending on the bacterial species.

Biological Role of Protein Glycosylation in Pathogenic Bacteria

Since most glycoproteins appear to be present in pili or flagella or are secreted into the environment, this location suggests that these proteins play a role in the interaction with the host. Glycosylation may contribute to the maintenance of protein conformation and stability including protection against proteolytic degradation, surface recognition, cell adhesion or immune evasion. For the AIDA-I adhesion of enterotoxic *E. coli* H10407 strain, a lack of glycosylation abolishes adherence to host cells. A similar function has been described for the outer membrane protein of *C. trachomatis* which is disabled in receptor binding if not glycosylated. The last example is *C. jejuni* where glycosylation is important in intestinal colonization.

I. Benz, M.A. Schmidt **2002**, Never say never again: protein glycosylation in pathogenic bacteria, *Mol. Microbiol.* 45, 267–276.

P. Messner **2004**, Prokaryotic glycoproteins: unexplored but important, *J. Bacteriol.* 186, 2517–2519.

6.5.4 Circular Polypeptides

Normally, polypeptides are linear chains of amino acids that fold into a three-dimensional conformation that defines their biological function. The termini of a polypeptide chain are often flexible and represent targets for proteases. Therefore, chemical linkage of the N and C termini in a peptide bond should stabilize proteins. Do circular proteins occur in nature? Indeed, depending on their biological function, two classes of circular proteins have been described in bacteria over the past few years: peptides and small proteins with antibacterial activity and cyclic pilins.

The largest known circular protein with antibacterial activity is bacteriocin AS-48 from *E. faecalis*. This highly basic 70-amino acid protein protects the producing strain against other bacteria by forming pores in the cytoplasmic membrane causing proton loss owing to permeabilization of the membrane. A second circular protein is termed microcin J25 which is secreted by strains of *E. coli*. This 21-amino acids highly hydrophobic peptide interferes with cell division. Both peptides are encoded by plasmids and are synthesized as longer precursors; and the genes required for maturation and secretion are also found on the plasmids.

The second class of circular proteins plays a role in bacterial conjugation. As outlined under 10.2, the initial step in bacterial conjugation requires physical contact between the donor and the recipient cell. This contact is produced by tube-like extracellular filamentous structures, the six pili which consist of pilin subunits.

Pilins encoded by RP4 and similar plasmids contain 73–78 amino acid residues on a continous backbone of peptide bonds. Cyclization of these polypeptides occurs by two different mechanisms, one exemplified by the pilin of RP4 encoded by the *trbC* gene and the other by the VirB2 pilin found in the T pilus of *A. tumefaciens*. The RP4 pilin is synthesized as a 145-residue prepro-TrbC which is processed in three distinct steps. First, an as-yet-unidentified protease removes 27 residues from the C-terminus. Next, the 36-residue signal peptide is cleaved off by the signal peptidase LepB. The third and last step is carried out by the TraF serine peptidase, which first removes four residues from the C-terminus and then is involved in the cyclization of remaining 78-residue pilin. In the case of the 121-amino-acid VirB2 pro-pilin, this polypeptide chain is processed into T-pilin of 74 amino acid residues by the signal peptidase (Fig. 6.52). The resulting T-pilin is coupled between the N-terminal Gln48 to Gly121 at the C-terminus in a head-to-tail peptide bond. This reaction is catalyzed by a peptide cyclase (Fig. 6.52).

M. Trabi, D.J. Craik **2002**, Circular proteins – no end in sight, *Trends Biochem. Sci.* 27, 132–138.

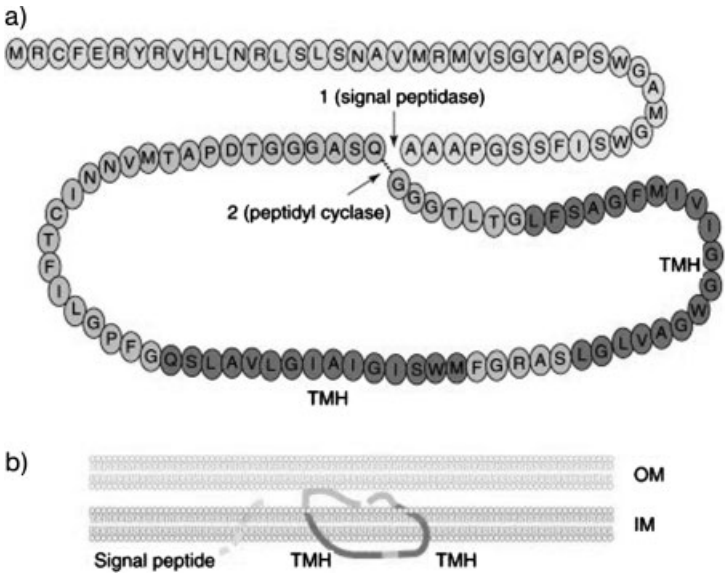


Fig. 6.52 Processing of the VirB2 pro- into the mature cyclic T-pilin by a peptidyl cyclase. (A) Amino acid sequence of the pro-pilin including the processing site. (B) Insertion of the pro-pilin into the cytoplasmic membrane and removal of the signal peptide. E. Lai, C.I. Kado **2000**, *Trends Microbiol.* 8, 361–369; Fig. 2.

7
Protein Quality Control Through Molecular Chaperones, Protein Folding Catalysts and ATP-dependent Proteases

7.1
Introduction

E. coli cells utilize up to 20000 ribosomes to produce an estimated total of 30000 polypeptides per minute. Each polypeptide chain has to fold into its unique three-dimensional native structure, starting from the unfolded linear chain. How this is achieved is one of the most interesting questions in molecular biology. When does the process of folding begin in the lifetime of a protein? Do proteins fold spontaneously or do they need help from other proteins? What happens to proteins which become (partially) unfolded due to stressful conditions or as a consequence of amino acid replacements due to mutations in their genes? Work carried out over the past 20 years in many laboratories around the world led to the discovery of *molecular chaperones* which aid protein folding, true protein folding catalysts (*prolyl-peptidyl isomerases* and *oxidoreductases* involved in disulfide bond formation) and *ATP-dependent proteases*. Molecular chaperones and ATP-dependent proteases monitor the folded state of other proteins. In addition to recognizing non-native conformations, these protein quality control factors distinguish substrates that can be refolded from those that need to be degraded. Molecular chaperones and ATP-dependent proteases constitute a network to ensure protein quality control within living cells.

7.2
Molecular Chaperones

The term ''molecular chaperones'' has been assigned to a broad range of protein families whose common property is that they recognize misfolded and denatured proteins (commonly denoted as non-native proteins) and protect the cell from the toxic and pathogenic consequences of protein aggregation. These non-native proteins are not only a byproduct of the folding process but are also produced through fluctuations in the native structure, the rate of which increases under extreme physicochemical conditions. Molecular chaperones bind transiently to non-native structures in polypeptide chains to prevent illegitimate protein–protein in-

Dynamics of the Bacterial Chromosome. Structure and Function. W. Schumann
Copyright © 2006 Wiley-VCH Verlag GmbH & Co. KGaA
ISBN 3-527-30496-7

teractions which may led to protein aggregates. The recognition of non-native proteins is mediated by exposed hydrophobic areas, the polypeptide backbone and/or specific secondary structures. Depending on their mode of action, four classes can be distinguished: *folder chaperones*, *holder chaperones*, *disaggregating chaperones* and *intramolecular chaperones*. Folder chaperones play an active role in folding and unfolding, using ATPase cycles to promote cycles of non-native protein binding and release by several mechanisms. In contrast, holder chaperones only bind to non-native proteins without further influencing their refolding. Disaggregating chaperones bind to protein aggregates and dissolve them into smaller aggregates which can be further worked on by the DnaK system. Intramolecular chaperones are part of proproteins, involved in their folding and then cleaved off. The minimal chaperone machinery found in the known bacterial genomes consist of the nascent chain binding chaperones trigger factor and DnaK (together with its cofactors DnaJ and GrpE) as well as the GroEL/GroES system that acts post-translationally in folding. A list of important molecular chaperones together with recognized target proteins is presented in Table 7.1.

Table 7.1 Important molecular chaperones and their target proteins.

Class of chaperone	*Name of molecular chaperone*	*Target protein(s)*
Folder chaperone	DnaK	Numerous
	GroEL	Numerous
	ClpA	e.g., RepA dimers
	ClpX	e.g., λO protein; MuA-DNA complexes
	ClpY	SulA
Holder chaperone	Trigger factor	Nascent chains emerging from ribosomes
	SecB	Many periplasmic proteins
	sHSPs	Non-native proteins
	Hsp31	Early unfolding intermediates
	Hsp33	Oxidatively damaged proteins
	PapD	Pilins
Disaggregating chaperone	ClpB	Protein aggregates
Intramolecular chaperone	pro	e.g., extracellular proteases

7.2.1 Folder Chaperones

Folder chaperones typically utilize ATP binding and hydrolysis to switch from a low- to a high-affinity binding state. Different classes of foldases, ATP-dependent molecular chaperones, have evolved and three different systems can be found in all eubacterial species, the DnaK, the GroE and the Clp systems, which will be described in detail.

The DnaK Chaperone Machine

The DnaK proteins (the eukaryotic homolog is termed Hsp70) constitute a large family of highly conserved chaperones that assist a multitude of protein-folding processes, including *de novo* folding of proteins, prevention of aggregation and refolding of stress-denatured proteins, disaggregation of protein aggregates, control of activity and stability of regulatory proteins and degradation of unfolded proteins. These diverse functions make DnaK outstanding among all known molecular chaperones.

The DnaK system consists of three different proteins, the chaperone DnaK (Hsp70) the co-chaperone DnaJ (Hsp40) and the nucleotide exchange factor GrpE, which act together and are often called the DnaK chaperone machine or team. DnaK consists of two functional domains: a ~44-kDa N-terminal ATPase domain with GrpE-binding affinity and a ~27-kDa C-terminal peptide-binding domain which binds polypeptide substrates and DnaJ (Fig. 7.1). It should be mentioned that DnaK contains another ~100 amino acid residues at its immediate C-terminal end of unknown function. The N-terminal domain forms two lobes with a deep cleft; and an ATP molecule binds to the bottom of the cleft. The peptide-binding domain consists of two subdomains, a β-sandwich subdomain with a peptide-binding cleft and an α-helical lid segment. Opening and closing of the lid is mediated via regulated changes in the ATPase domain of the protein. Target peptides are about seven residues long and are typically hydrophobic in their central region, with isoleucine and leucine being preferred. These binding sites occur statistically every 40 residues in polypeptide chains and are recognized with affinities of 5 nM to 5 μM. Target regions are bound to DnaK in an extended state through hydrophobic side-chain interactions and hydrogen bonds with the peptide backbone.

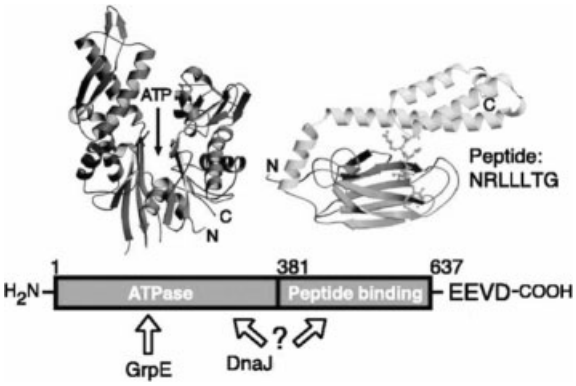


Fig. 7.1 Domain organization of the DnaK chaperone. DnaK consists of an N-terminal ATPase and a C-terminal peptide-binding domain. The ATPase domain consists of two lobes which form a deep cleft serving as a binding pocket for ATP. The peptide-binding domain (here shown with a heptapeptide substrate) forms a β sandwich composed of two sheets of four strands each, followed by α helices spanning back over the sandwich. Further indicated by open arrows are interaction sites for GrpE and DnaJ. F.U. Hartl, M. Hayer-Hartl **2002**, *Science* 295, 1852–1858; Fig. 3A. (This figure also appears with the color plates.)

DnaK possesses a weak intrinsic ATPase activity that is regulated by a co-chaperone belonging to the so-called J-proteins, which are characterized by an evolutionary conserved motif of ~75 amino acid residues known as the J-domain. Apart from the presence of a J-domain, the J-proteins are diverse. The major co-chaperone of DnaK, DnaJ, in addition to stimulating the ATPase activity of DnaK, possesses chaperone activity as revealed by its capacity to recognize non-native proteins and prevent the aggregation of folding intermediates *in vitro*. *E. coli* DnaJ is a homodimer, and each 41-kDa subunit is composed of four successive domains from the N-terminus: the conserved J-domain responsible for stimulating the ATPase activity of DnaK, a glycine-rich region (G/F) in conjugation with the J-domain for productive interaction with DnaK, a zinc finger-like domain and a less conserved C-terminal domain for substrate binding. The third player of this team, the nucleotide exchange factor GrpE, a homodimer of 20-kDa subunits, consists of 34 N-terminal residues of unknown function, followed by an α-helical domain involved in dimerization and a β-sheet domain. Both the central α-helical and the C-terminal β-sheet domain are needed for DnaK binding.

The DnaK chaperone is prone to bind a non-native substrate protein with a bound ATP molecule and the α-helical lid open (Fig. 7.2). Substrate proteins either bind directly to the peptide-binding domain of DnaK or, alternatively, are targeted by the co-chaperone DnaJ to DnaK. Rapid substrate binding occurs in the ATP-bound state of DnaK in which the lid is in the open conformation. In the ATP-bound state, substrate molecules bind with low affinity and fast exchange rates since the pocket stays open. Stable binding of the substrate protein involves

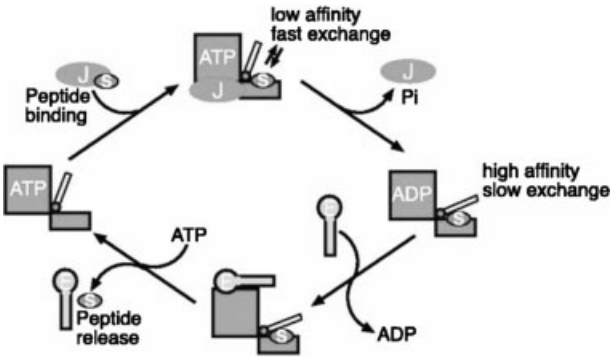


Fig. 7.2 Reaction cycle of the DnaK chaperone machine. DnaK-ATP with the open lid constitutes the active form able to accept non-native substrate polypeptide chains. These are either targeted to DnaK by DnaJ (as shown here) or bind directly to the peptide-binding domain of DnaK. The initial binding of the substrate protein is characterized by low affinity allowing fast dissociation. Binding of DnaJ to DnaK stimulates its ATPase activity resulting in the release of inorganic phosphate followed by closing of the lid, as shown in the cartoon. Now, the substrate protein is trapped and can start to refold. Later, GrpE binds to the complex, stimulates release of ADP followed by opening of the lid with concomitant dissociation of the substrate protein. After binding of ATP, a new reaction cycle can be started. F.U. Hartl, M. Hayer-Hartl **2002**, *Science* 295, 1852–1858; Fig. 3B.(This figure also appears with the color plates.)

closing of the lid triggered by hydrolysis of the bound ATP molecule to ADP and inorganic phosphate through DnaJ. DnaJ interacts with DnaK through a His-Pro-Asp motif in DnaJ. Hydrolysis of ATP is the rate-limiting step in the ATPase cycle of DnaK. The ADP-bound state of DnaK is characterized by high affinity binding and slow exchange rates for substrates. The cycling of DnaK between these two states is regulated by DnaJ and by GrpE. The N-terminal J-domain of DnaJ binds to DnaK thereby facilitating peptide binding. GrpE induces the release of ADP from DnaK, and upon rebinding of ATP, the DnaK–substrate complex dissociates. Two salt bridges (Lys55-Glu267, Arg56-Glu264) and one exposed loop consisting of Ala276-Arg302 are important for DnaK to interact with GrpE.

DnaK also promotes the disassembly of oligomeric protein complexes. This has been shown to occur with phage P1 RepA protein, which can be isolated as an inactive homodimer. This dimer is recognized by DnaJ and the DnaJ-RepA dimer then interacts with DnaK with subsequent conversion of the dimer to active monomers. In a similar way, an inactive complex consisting of the DnaB helicase and phage λ O and P proteins is disassembled by DnaK and DnaJ, leading to the activation of helicase activity of DnaB. Recently, it could be shown that the DnaK chaperone team together with ClpB can disaggregate any protein aggregate (see below).

In *E. coli*, two additional DnaK homologs have been identified, *h*eat-*s*hock *c*ognate HscA (Hsc66) and HscC. HscA exhibits about 40% sequence identity with DnaK and assists in the biogenesis of iron–sulfur proteins. Hsc66 interacts with the IscU, a small, dimeric protein that is capable of forming labile iron–sulfur clusters *in vitro* and is proposed to serve as a template for iron–sulfur cluster formation *in vivo*. HscA binding of IscU may serve to regulate the type or stability of iron–sulfur clusters formed on IscU and this interaction is enhanced by the J-domain of the co-chaperone Hsc20, which directly binds to both IscU and HscA. HscC shares about 30% amino acid identity with DnaK and HscA and therefore seems to be a more distant relative of DnaK. HscC does not act as a general chaperone, since it does not assist refolding of a denatured model substrate proteins. Its ATPase activity is specifically stimulated by Hsc56, a 56-kDa protein with a J-domain. *E. coli* cells with a deletion of *hscC* revealed a hypersensitivity to Cd^{2+} ions and UV irradiation, suggesting roles of HscC in the cellular responses to this stress regimen.

The GroE Chaperonin Machine

The chaperonins are a conserved class of large double-ring complexes of ~800 kDa enclosing a central cavity. They are present in nearly all organisms and are essential for viability, because they are required for some essential proteins. They occur in two subgroups with a similar architecture but distantly related in amino acid sequence:

1. Group I contains the eubacterial chaperonins termed GroEL and its co-chaperonin GroES and homologous proteins from mitochondria and chloroplasts.
2. Group II includes the yeast chaperonin CCT, present in the cytosol of eukaryotes and archaea.

While the GroEL chaperonins are rather promiscuous and bind most non-native proteins *in vitro*, the CCT chaperonins are more selective by being involved in the folding of the cytoskeletal proteins actin and tubulin.

GroEL (also known as Hps60), the best characterized chaperonin, is a homooligomeric complex of 58 kDa subunits arranged in two seven-membered rings stacked back to back. Each monomer consists of three functional domains: a mobile *apical domain* responsible for binding both substrate proteins and GroES, a small *intermediate hinge domain* and an *equatorial domain* containing the ATP binding sites and most of the intersubunit contacts (Fig. 7.3a). The 14-mer consists of two stacked rings of seven subunits each; and the whole assembly is held together by the back-to-back rings of the 14 equatorial domains that form most of the intra- and all of the inter-ring contacts (Fig. 7.3b). The co-chaperonin GroES also forms a ring-like structure consisting of seven identical 10-kDa subunits, and this GroES ring interacts with the apical domains the GroEL barrel-shaped complex through mobile loops thereby closing one end like a lid (Fig. 7.3c). Together with GroES, GroEL captures, encapsulates and releases its substrates in cycles driven by ATP binding and hydrolysis.

The GroEL ATPase and folding cycle is presented in Fig. 7.4. A non-native protein makes contact with hydrophobic binding sites of the *cis*-ring that line a cavity at either end of the barrel-shaped complex. Next, the binding of ATP to the equatorial domains of each GroEL subunit initiates large conformational changes. The GroEL apical domains swing 60° upwards and twist through 90° and are now able to contact the mobile loops of the GroES lid. This conformational change buries the hydrophobic patches in the wall of the folding chamber, which leads to the release of the non-native polypeptide chain into the cavity, where it starts to fold. Furthermore, the space within the cavity is enlarged. Next, the ATP molecules are hydrolyzed; and the binding of ATP to the subunits of the *trans*-ring signals the dissociation of the GroES lid and release of the substrate protein from the *cis*-ring. At about the same time, another non-native polypeptide chain binds to the hydro-

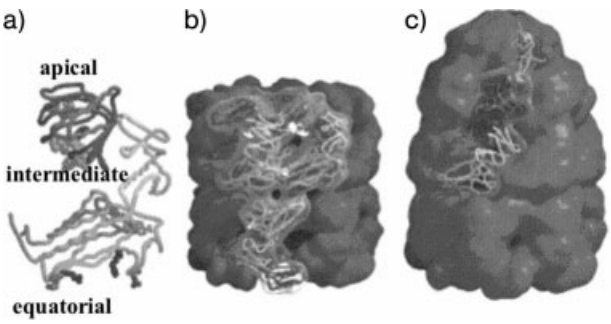


Fig. 7.3 The GroE chaperonin complex. (a) The three functional domains of the GroEL monomer. (b) Structure of two GroEL rings (14-mer) with three subunits shown at the front. (c) The GroEL-ADP-GroES complex with one GroEL and GroES subunit each. H.R. Saibil, N.A. Ranson **2002**, Trends Biochem. Sci. 27, 627–632; Fig. 2. (This figure also appears with the color plates.)

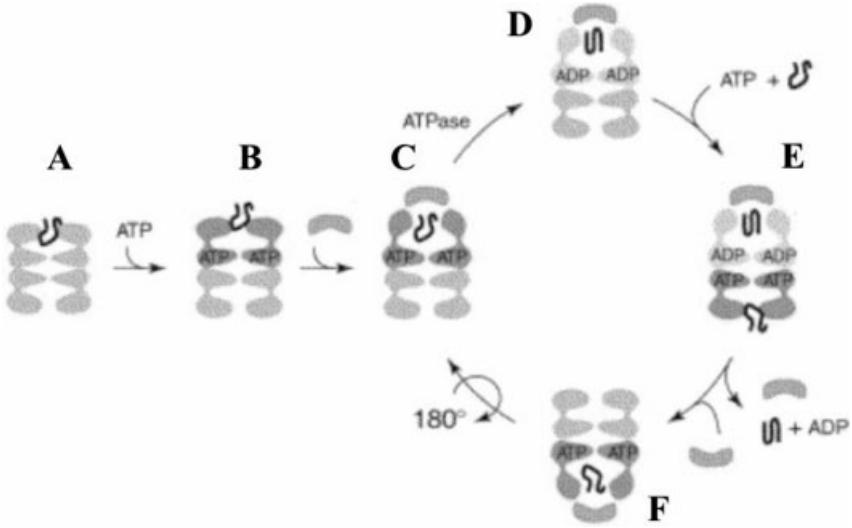


Fig. 7.4 The GroEL-GroES reaction cycle. (A) A substrate protein diffuses into the upper ring where it will bind to hydrophobic side chains exposed on the inner surface of the apical domain. (B) Next, each equatorial domain of the upper ring binds one molecule of ATP with positive cooperativity. (C) This in turn triggers binding of a GroES ring to the upper ring causing an upward movement of the apical domain, retraction of the hydrophobic side-chains and release of the substrate protein into the cavity of the upper GroEL ring where it starts to fold. (D) ATP is hydrolyzed and (E) another substrate molecule diffuses into the lower ring, followed by binding of ATP. (F) GroES diffuses away from the upper ring, followed by the substrate protein; and the substrate molecule bound to the lower ring gets encapsulated. H.R. Saibil, N.A. Ranson **2002**, Trends Biochem. Sci. 27, 627–632; Fig. 1. (This figure also appears with the color plates.)

phobic binding sites of the *trans*-ring, the ADP from the *cis*-ring is released and another folding cycle starts. It takes about 15 s from the binding of a non-native protein to GroEL to its release, as measured by *in vitro* experiments. This period of time may not be sufficient to allow refolding of a substrate protein. Therefore, after release, a protein molecule not completely folded will bind either to the same or to another GroEL chaperonin machine.

ATP hydrolysis acts as a timer, giving the substrate ~10 s to fold before GroEL is primed to release GroES. In *E. coli*, a wide range (~10%) of newly synthesized polypeptides interacts with GroEL, although it is not known how many of those proteins really depend on the chaperonin for folding. This question is illustrated by the observation that GroEL can bind and release actin and tubulin in an ATP-dependent manner, but it cannot promote their folding. There are also specialized co-chaperonins: bacteriophages T4 and RB49 enocode their own GroES homologs (GP31 and CocO, respectively) that cooperate with GroEL to fold certain phage capsid proteins whose folding and assembly are not supported by the endogenous GroES. These GroES homologs cause significant changes in cavity size and chemical environment.

The manner in which the GroEL structure change promotes protein folding remains elusive. Two models, not mutually exclusive, have been proposed, the *Anfinsen cage* and the *iterative annealing model*. The Anfinsen cage model is based on the view that the GroEL cavity provides a sequestered microenvironment where folding to the native state can proceeed. But the substrate protein is ejected from the cavity with each round of ATP hydrolysis, independent of having reached its native state or not. The iterative annealing model assumes that the rate-limiting step in slow protein folding is the intramolecular reorganization of misfolded and trapped protein segments. This model suggests that ATP hydrolysis is coupled to unfolding of the non-native substrate protein, so that the misfolding is relieved and forward folding can resume either within the protected cavity or outside GroEL.

The Clp Chaperones

The Clp (Hsp100) ATPases are a unique group of ATP-dependent chaperones associated with the disassembly of protein complexes. These enzymes have, therefore, become known as "*protein remodeling*" chaperones. The term protein remodeling has been used to describe many different reactions that result in a change in the function, and often conformation, of a target protein. The Clp ATPases are members of the AAA family of proteins; and the key feature of this family is a highly conserved AAA module of about 230 amino acids that is present in one or two copies in each protein. Each module can typically be divided into two subdomains, designated αβ and α. In the αβ subdomain, there is a conserved Walker A motif, involved in binding the phosphate of ATP, and a Walker B motif, involved in metal binding and ATP cleavage. A unifying structural feature of these AAA proteins is the arrangement of the subunits into ring-shaped hexameric or heptameric complexes. Most Clp chaperones also form complexes with proteases and participate directly in ATP-dependent protein degradation (see below). Depending on the number of ATP-binding domains, the Clp chaperones can be divided into two families (Fig. 7.5):

- The Hsp100 family contains two distinct ATP-binding domains (ClpA, ClpB, ClpC, ClpD).
- The second family carries only one ATP-binding domain (ClpM, ClpN, ClpX, ClpY).

All Clp chaperones form oligomeric ring-like structures and recognize specific proteins, then promote a conformational change in these proteins using the energy of ATP hydrolysis. These Clp-catalyzed changes in protein conformation have many biological consequences, such as protein remodeling reactions in which multimeric protein complexes are dissembled into their component subunits, exhibiting biological activity.

The *ClpA* ATPase component forms hexamers in the presence of Mg-ATP. One of the substrate proteins recognized by ClpA is the RepA dimer encoded by the P1 phage. These dimers bind ClpA hexamers in a reaction requiring ATP but not ATP hydrolysis. Upon ATP hydrolysis, RepA is released from ClpA and simulta-

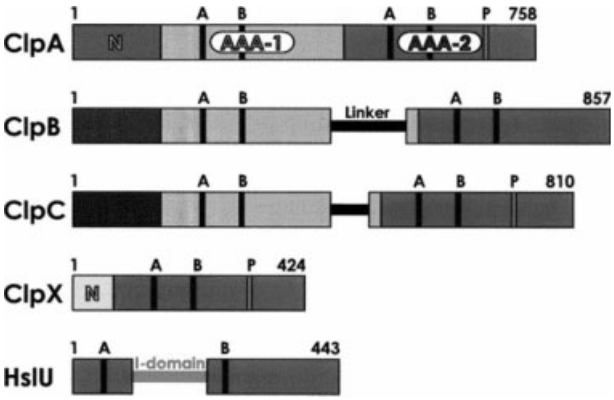


Fig. 7.5 Structural domains of AAA$^+$ proteins. The ATPase domains (AAA-1, AAA-2) contain the Walker A (GX4GKT) acid and Walker B (HyDE) nucleotide; P, peptidase recognition motif; the I-domain is proposed to mediate substrate interaction. D.A. Dougan, et al. **2002**, *FEBS Lett.* 529, 6–10; Fig. 1.

neously converted into the active monomer form. This reaction can also be carried out by the DnaK chaperone team (see above).

ClpX consists of an N-terminal zinc-binding domain (ZBD) followed by one AAA ATPase domain. ClpX prevents heat-induced aggregation of phage λO protein and is able to disaggregate heat-denatured λO in an ATP-dependent reaction. Furthermore, ClpX participates in the DNA replication of phage Mu in disassembling MuA-DNA complexes. MuA is a monomeric protein that forms stable tetramers upon binding to specific Mu DNA sequences during Mu DNA transposition. The recognition motif is located within the C-terminal ten amino acids. This signal can be transferred to a different protein, which is then recognized by ClpX. ClpX also dissociates inactive dimers of the initiator protein of DNA replication of the plasmid RK2 into active monomers.

The *ClpY* (*HslU*) 19-kDa chaperone prevents aggregation of the cell division inhibitor SulA in a concentration-dependent fashion. While the oligomeric form of ClpY is needed to inhibit aggregation, ATP-binding is not required.

A. Wawrzynow, et al. **1996**, The Clp ATPases define a novel class of molecular chaperones, *Mol. Microbiol.* 21, 895–899.

7.2.2
Holder Chaperones

In contrast to folder chaperones, holder chaperones are usually ATP-independent chaperones that bind non-native proteins when all the folder chaperones are loaded with denatured proteins, to prevent the formation of aggregates. When folder chaperones have released their polypeptide chains, holder chaperones deliver their cargo to them. Alternatively, holder chaperones can also pass non-native proteins into the proteolytic system.

The Trigger Factor

Proteins contain within their complete amino acid sequence all of the information necessary for attaining their functional three-dimensional structure. But all newly synthesized proteins face challenges in reaching their native state within the crowded environment of the cell. While some domains of a nascent chain might be capable of folding spontaneously, the folded structure cannot be obtained until the entire domain is synthesized. This time lag increases the chance that hydrophobic sequences normally buried in the interior of the protein will become exposed, resulting in protein aggregation. About 40 amino acids of the nascent chain are protected from the cytosol by the ribosome exit tunnel. When the chain leaves the tunnel, molecular chaperones bind preventing aggregation.

Bacteria have a specialized molecular chaperone called trigger factor (TF), a 48-kDa protein, with both peptidyl-prolyl *cis*/*trans* isomerase (PPIase) activity and a holder chaperone-like function. TF is a dimer in solution and consists of three domains. The N-terminal domain (aa 1–144) mediates ribosome binding, while the central domain (P domain, aa 145–247) has PPIase and substrate-binding activity. The function of the C-terminal domain (C domain, aa 248–432) is unknown. About half the cellular level of TF in *E. coli* is bound to the large ribosomal subunit in a stoichiometry of approximately 1:1 (Fig. 7.6a). Crosslinks were obtained between L23 and L29 (both ribosomal proteins are located at the exit tunnel), where the interaction with L23 seems to be the critial one. Substitution of specific residues of L23, exposed on the ribosome surface, disrupts TF binding to the ribosome. TF can also be crosslinked to short nascent chains of 56 amino acids, indicating that this chaperone binds to polypeptides when they exit the ribosome, while the other half is free in the cytoplasm. TF associates cotranslationally early with most nascent polypeptides of cytosolic and secreted proteins independently of proline residues. It was therefore proposed that TF is the first chaperone that interacts with nascent chains and assists cotranslational protein folding. Genetic evidence indicates that TF cooperates with the DnaK system (Fig. 7.6b) to ensure proper folding of cytosolic proteins. While the lack of TF or DnaK alone has almost no effect on cell growth between 30 °C and 37 °C, the absence of both TF and DnaK is lethal. This synthetic lethality can be overcome by enhanced synthesis of another holder chaperone, SecB (see below).

E.A. Craig, et al. **2003**, Ribosome-tethered molecular chaperones: the first line of defense against protein misfolding? *Curr. Opin. Microbiol.* 6, 157–162.

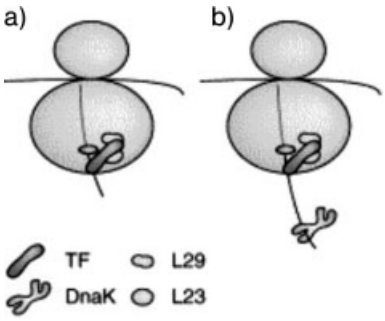


Fig. 7.6 Protein folding of nascent polypeptide chains. (a) Trigger factor binds to the exit tunnel of the ribosome by interacting with ribosomal proteins L23 and L29. It interacts with the nascent polypeptide chain when it emerges at the exit tunnel. (b) DnaK binds to the nascent chain if it extrudes from the ribosome. Both molecular chaperones prevent premature folding of the nascent polypeptide chain. E.A. Craig **2003**, *Curr. Opin. Microbiol.* 6, 157–162; Fig. 1.

The SecB Protein

SecB is a cytosolic protein important, but not essential, for the translocation of proteins to the periplasm in many Gram-negative bacteria. SecB is a highly acidic homotetramer with a subunit molecular weight of ~17 kDa which has two functions:

- It serves to keep the preprotein in an unfolded, translocation-competent state (*anti-folding activity*).
- It facilitates targeting of the preprotein to membrane-associated SecA (*targeting activity*).

SecB associates with ribosome-bound nascent chains of a subset of periplasmic and outer membrane proteins after they have reached a length of ~150 residues. The typical SecB-binding motif is approximately nine residues long and enriched in aromatic and basic residues; and the majority of the binding energy for these peptides results from hydrophobic interactions. Binding of SecB to these preproteins prevents premature folding and aggregation and also pilots the preproteins to the peripheral translocon component SecA. Recently, it has been shown that SecB can interact with short nascent chains of cytoplasmic proteins, too, demonstrating that SecB acts as a generalized chaperone. SecB efficiently suppresses both the temperature sensitivity and the aggregation-prone phenotype of a strain lacking both TF and DnaK. The simultaneous deletion of the *dnaK* and *tig* genes in *E. coli* causes synthetic lethality, but cells can tolerate the deletion of both chaperones at temperatures <30 °C. SecB can interact both co- and post-translationally with short nascent chains of both secretory and cytosolic proteins, suggesting that SecB acts as a *bona fide* generalized chaperone (see above).

Small Heat-shock Proteins

The family of α-*h*eat-shock *p*roteins (α-HSPs, sometimes also called small heat-shock proteins) occurs in all organisms and is characterized by the following features:

1. Their size range is 12–43 kDa, with the majority being between 14 kDa and 27 kDa.
2. They form oligomeric, globular structures consisting of 4–42 subunits. The number of subunits and the degree of flexibility in the complex varies depending on the species. While Hsp16.3 from *M. tuberculosis* first trimerizes and then oligomerizes into 9-mers (Fig. 7.7a), *M. jannaschii* Hsp16.5 forms dimers which subsequently build a stable hollow sphere with openings of 24 subunits (Fig. 7.7b).
3. While the overall homology among α-HSPs is rather low, they exhibit stretches of sequence homology, mostly in the C-terminal part of the protein, designated the α-crystallin domain. This name is derived from the α-crystallin chaperones in the eye lens, where they are supposed to prevent the early aggregation and precipitation of lens proteins.

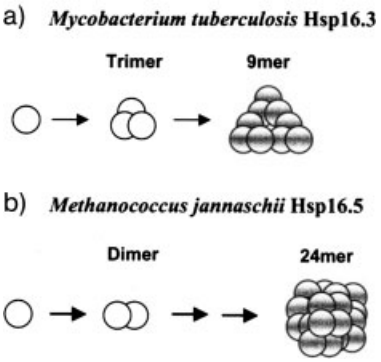


Fig. 7.7 Small heat-shock proteins. (a) *M. tuberculosis* Hsp16.3 first trimerize and then form 9-mers. (b) *M. jannaschii* Hsp16.5 dimerize and then build up 24-mers. F. Narberhaus **2002**, *Microbiol. Mol. Biol. Rev.* 66, 64–93; Fig. 4.

4. α-HSPs are generally believed to be ATP-independent chaperones.
5. The major function of α-HSPs within cells is to bind non-native proteins, to prevent their aggregation.

α-HSPs are rather promiscous in binding all kinds of non-native proteins. What is the fate of these bound proteins? There seem to be three possibilities. Most polypeptide chains are handed over to foldases involved in refolding. Some proteins may be caught by ATP-dependent proteases that degrade them to small peptides 5–15 residues in length. A third possibility is spontaneous release. These polypeptides may refold without the aid of a foldase, or they may become bound by a foldase or by a protease. While some α-HSPs are already present at high amounts at physiologial temperatures, others are clearly heat-inducible: e.g., the two α-HSPs of *E. coli*, where the genes *ibpA* and *ibpB* are about 300-fold induced after a heat shock from 30 °C to 42 °C. Most interestingly, these two α-HSPs were first identified as components of inclusion bodies which are often formed upon high-level production of recombinant proteins in *E. coli*. Therefore, the designation of these two genes was derived from *i*nclusion *b*ody *p*roteins.

The Hsp31 Chaperone

Hsp31, the product of the *hchA* (*h*eat-inducible *ch*aperone *A*) gene, is a heat-inducible homodimeric protein of 31 kDa. It is highly conserved in a number of pathogenic eubacteria and fungi and relies on temperature-induced conformational changes to expose structured hydrophobic domains to the solvent that likely serve as binding sites for partially folded substrate proteins. In contrast to the small heat-shock proteins which do not release their substrates upon temperature downshift, a significant fraction of client proteins is rapidly ejected from the holder chaperone in an active form when binary complexes are transferred from 45 to 23°C. This reaction could be triggered by binding of ATP the concentration of which is low at high temperatures but increases when cells return to physiological conditions. These results indicate that Hsp31 captures and transiently stabilizes early unfolding intermediates to alleviate overloading of the DnaK-DnaJ-GrpE machin-

ery under conditions of severe thermal stress. Loss of Hsp31 in *E. coli* has a strong deleterious effect on the ability of cells to survive and recover from transient exposure to 50 °C and leads to the enhanced aggregation of a subset of host proteins at this temperature. Hsp31 lacks an ATPase activity which is necessary to fuel the active remodeling of protein aggregates.

The Redox-regulated Molecular Chaperone Hsp33

Hsp33 is a highly conserved heat-shock protein which is present in the cytoplasm of more than 50 different eubacterial species and belongs to a novel class of redox-regulated proteins, whose activity is regulated by their redox state. The Hsp33 gene in *E. coli* is under the control of σ^{32} (see Section 9.2); and the steady-state level of Hsp33 increases to ~3 μM, twice the level of Hsp33 under nonstress conditions. On a post-translational level, Hsp33 is under oxidative stress control, enabling the chaperone to sense changes in the redox conditions of the environment. The redox sensor is located in the C-terminus of the protein and consists of a novel, very high affinity zinc-binding motif ($CXCX^{27-32}CXXC$) in *E. coli*. The four absolutely conserved cysteine residues constitute the redox switch and are kept in the reduced deprotonated thiolate anion state and together coordinate one zinc(II) ion (K_d ~10^{-18} M; Fig. 7.8). Under oxidative conditions, these four cysteines release zinc and rapidly form two intramolecular disulfide bonds, connecting the two pairs of neighboring cysteines, Cys^{232} with Cys^{234} and Cys^{265} with Cys^{268}. Full activation of the Hsp33 chaperone is a two-step process. While the first step results in the formation of the two intramolecular disulfide bonds and

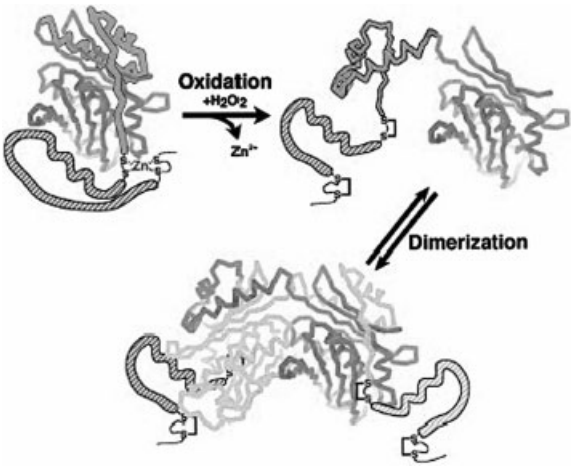


Fig. 7.8 Model of Hsp33 activation. Under reducing conditions, Hsp33 occurs in a monomeric form and four cysteine residues are involved in zinc coordination. After addition of H_2O_2, zinc is released and two intramolecular disulfide bonds are formed. Next, two oxidized Hsp33 monomers form a highly active dimer. P.C.F. Graf, U. Jacob **2002**, *Cell. Mol. Life Sci.* 59, 1624–1631; Fig. 1. (This figure also appears with the color plates.)

concurrent release of zinc accompanied by major conformational rearrangements in the protein, the second step involves the dimerization of two oxidized Hsp33 monomers (Fig. 7.8). This second step is temperature- and concentration-dependent; and oxidized Hsp33 dimers are very potent molecular holder chaperones, which prevent the aggregation of a variety of unfolded proteins. To prime Hsp33 for substrate protein release, reducing conditions have to be restored. Reduction is catalyzed by the glutaredoxin and thioredoxin systems *in vivo* and leads to the formation of highly active, reduced Hsp33 dimers. The reduced dimer is still active but now able to transfer the substrate proteins to the DnaK/DnaJ/GrpE folder system for refolding. Upon substrate release, reduced Hsp33 dimers dissociate into inactive monomers.

What is the *in vivo* function of Hsp33? Under reducing conditions, Hsp33 is predominantly reduced and therefore in the inactive state. If the two thiol-reducing systems present in *E. coli*, the thioredoxin and the glutathione pathways, are inactivated, cells accumulate disulfide-bonded proteins in the cytoplasm. Under these conditions, up to 60% of Hsp33 accumulates in the disulfide-bonded, active conformation; and these cells show a significantly higher resistance to H_2O_2 treatment. In contrast, deletion of the Hsp33 gene in these H_2O_2-stressed cells leads to a 10000-fold increase in H_2O_2 sensitivity. These observations led to the conclusion that activation of Hsp33 represents a first line of defense strategy against oxidative protein damage. Under conditions of severe oxidative stress at elevated temperatures, non-native proteins are no longer protected by the DnaK chaperone team, but bind to Hsp33. Here, the dramatic decrease in intracellular ATP levels that occur upon the exposure of cells to reactive oxygen species renders the ATPase domain of DnaK nucleotide depleted and thermolabile. Upon return to nonstress conditions, Hsp33 becomes inactivated, while DnaK resumes its active conformation.

Chaperones Involved in Pilin Assembly

Pili are fibrous organelles on the surface of Gram-negative cells that mediate attachment to surfaces (see Section 1.7). While some pili consist only of one subunit (the F pilus), others are composed of several different subunits which are assembled by various pathways. Here, I will focus on the chaperone–usher pathway as examplified by the Pap (P) pilus encoded by the 11 genes of the *pap* (*papA*–*papK*) locus. The *pap* locus (for *p*yelonephritis-*a*ssociated *p*ilus) is found in many uropathogenic strains of *E. coli* and codes for a thick rod with a thinner tip fibrillum at its distal end (Fig. 7.9). The rod is composed mainly of PapA subunits wound in a right-handed helix, while the tip fibrillum consists primarily of PapE subunits and ends with the PapG *adhesin*, which binds to gal(α1→4)gal-containing sugars found in the human kidney. PapE and PapG are linked by the two adaptor proteins PapF and PapK.

The biogenesis of the P pilus requires two dedicated proteins, the periplasmic chaperone PapD and the outer membrane protein PapC acting as an *usher*. PapD interacts with the pilus subunits individually (Fig. 7.9), facilitating its release from

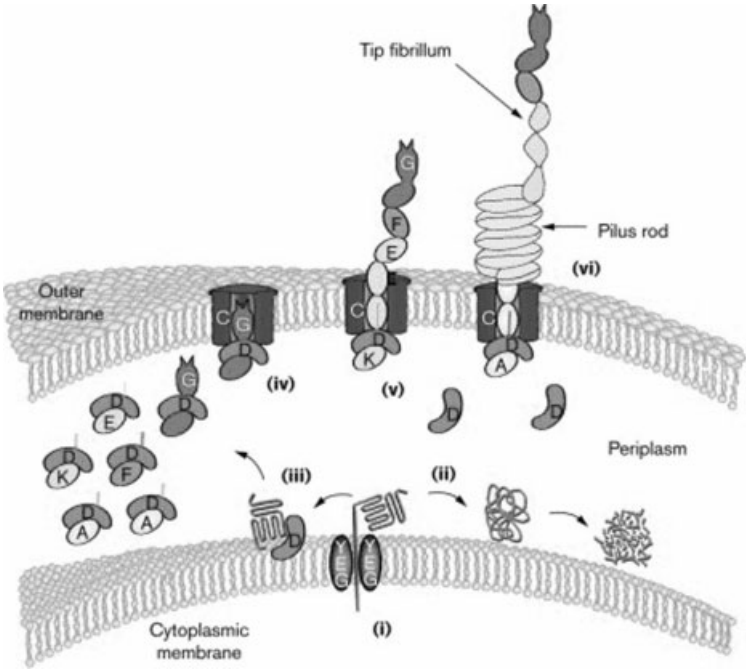


Fig. 7.9 Biogenesis of the P pilus. (i) The components of the P pilus and those involved in its biogenesis are synthesized in the cytoplasm and secreted into the periplasm using the Sec pathway. (ii) If components of the P pilus do not interact with the PapD chaperone, they fold inappropriately and are degraded by periplasmic proteases. (iii) Most P pilus components are bound by the PapD chaperone and guided to the PapC protein, an integral outer membrane protein acting as an usher (iv). (v) PapC monitors the correct assembly of the different components which form the active P pilus (vi). F.G Sauer, et al. **2000**, *Curr. Opin. Struct. Biol.* 10, 548–556; Fig. 1. (This figure also appears with the color plates.)

the cytoplasmic membrane, stabilizing the subunit and preventing premature subunit–subunit interaction in the periplasm. Subunits which escape the PapD chaperone are degraded. Next, the PapD–subunit complexes are targeted to the PapC usher forming a channel 2–3 nm in diameter, large enough to allow passage of the different subunits. The usher is thought to recognize the different subunits in the correct sequence and to facilitate dissociation of the subunits from PapD. The specificity of pilin selection into the growing pilus seems to be determined by the interaction of the hydrophobic groove of the pilin subunit complexed with the usher and the structure of the N-terminus in subsequent pilin subunits.

TTS Chaperones

The type III secretion (TTS) system is used by many pathogenic and symbiotic Gram-negative bacteria to deliver proteins within or beyond the membranes of

host cells (see Section 8.6.3). In short, TTS systems consist of approximately 15 proteins to construct the TTS apparatus called *injectisome*, two translocators forming a pore in the outer membrane, effector proteins that are injected into the eukaryotic cell and specific chaperones, the TTS chaperones. Genes coding for TTS chaperones are usually encoded by pathogenicity islands or plasmids which also inherit the remaining genes of the TTS system. TTS chaperones from different bacterial species do not exhibit sequence similarities, but they share common features, including a small size (about 15 kDa), an acidic pI and a predicted amphiphilic α-helix in their C-terminal parts. Usually, they act as dimers and bind to the N-terminal region of the cognate protein(s). TTS chaperones have been assigned different functions and may act as:

- antiaggregation and stabilizing factors
- signals for secretion
- antifolding factors
- regulators of gene expression.

In the first example, the TTS chaperone acts as an antiaggregation factor. The Ysc TTS system of *Y. enterocolitica* involves the two translocator proteins YopB and YopD (*Yersinia* *o*uter *p*roteins) and the cytoplasmic TTS chaperone SycD (for *s*pecific *Y*op *c*haperone *D*). SycD binds individually to both translocator proteins to prevent their aggregation within the cytoplasm. In the second example, another TTS chaperone of *Y. enterocolitica*, SycE, is involved in binding to the three effector molecules YopE, YopH and YopT, to target them to the secretion pathway. In the case of YopE, the N-terminal 130 amino acid residues constitute a functional SycE binding domain. In the third example, the TTS chaperone prevents folding of effector or/and translocator proteins, both of which have to travel through the injectisome. The secretion channel within the injectisomes is between 2 nm and 3 nm, allowing small proteins to travel folded, while large proteins have to be kept in their unfolded (secretion-competent) state. The last example involves the TTS chaperone as a regulator of gene expression (Fig. 7.10). As long as the bacterial cell does not make contact with its eukaryotic target cell, the substrate proteins to be secreted through the injectisome are kept bound by the chaperone. These substrate proteins are present at a low level, due to the largely inactive transcriptional activator protein. Upon contact with the eukaryotic cell, the substrate proteins are released from the chaperone and secreted, while the free TTS chaperone is able to bind to the transcription factor to convert it in its active form. The complex consisting of the active transcriptional activator and chaperone binds to the DNA upstream of the genes coding for the TTS genes to activate their transcription. One example is the SicA chaperone of *Salmonella* interacting with InvF. The complex SicA-InvF activates the transcription of a subset of genes involved in the TTS systems.

C. Parsot, et al. **2003**, The various and varying roles of specific chaperones in type III secretion systems, *Curr. Opin. Microbiol.* 6, 1–8.

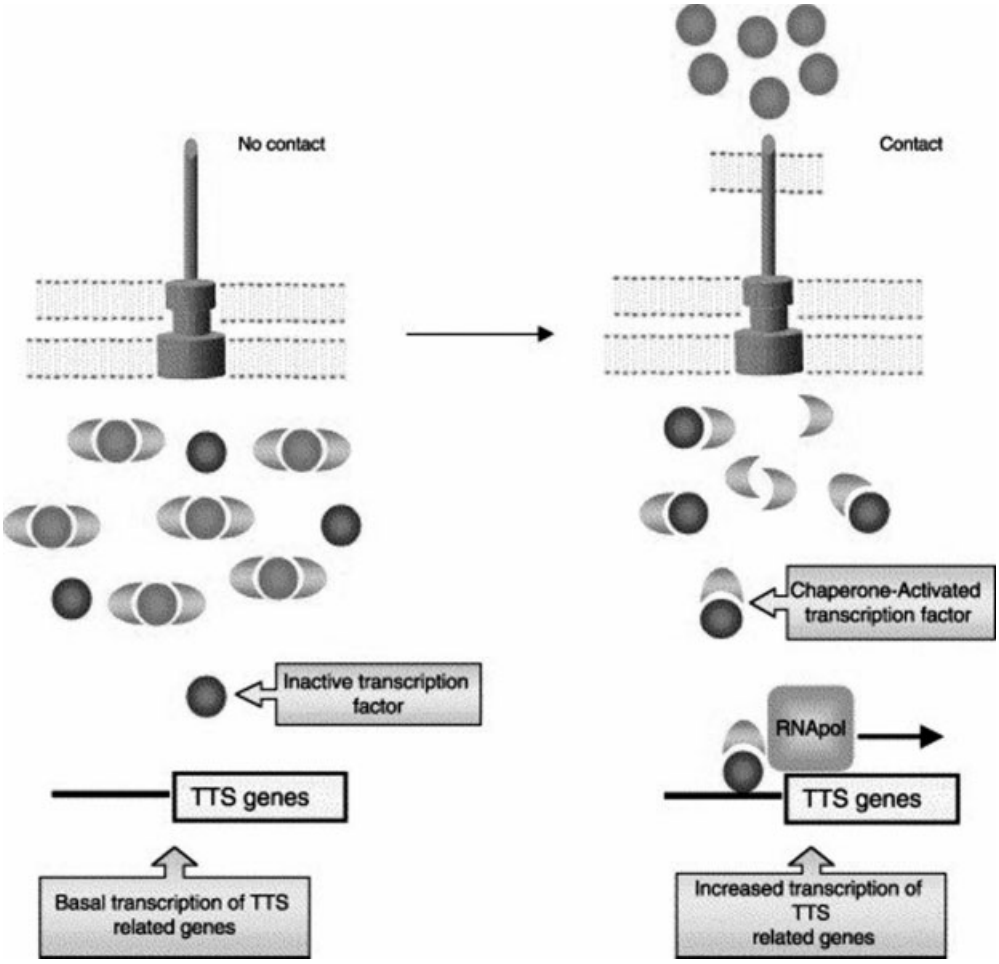


Fig. 7.10 TTS chaperone acts as a regulator of gene expression. In the absence of contact with a eukaryotic target cell, the dimeric TTS chaperone binds to the proteins to be secreted, and the transcriptional activator molecules present within the cytoplasm are inactive. Upon contact with a eukaryotic cell, the chaperone releases the substrate proteins, forms a complex with the transcription factor which is now able to interact with the DNA to stimulate transcription. M.F. Feldman, G.R. Cornelis **2003**, *FEMS Microbiol. Lett.* 219, 151–158; Fig. 4.

7.2.3 Disaggregating Chaperones

When high levels of unfolded proteins accumulate, the buffering capacity of their molecular chaperones and ATP-dependent proteases may be exceeded and proteins start to aggregate. Protein aggregation was once viewed as a dead end in the lifecycle of a protein. In 1990, S. Lindquist discovered a new heat-shock protein of

the yeast *S. cerevisiae*, termed Hsp104. The substrates of Hsp104 are protein aggregates generated, e.g., during severe heat stress. The *E. coli* ClpB is the homolog of the yeast Hsp104. Both Hsp104 and ClpB exhibit considerable sequence homology in their AAA^+ domains, which are important for ATP binding, hydrolysis and oligomerization. Each ClpB subunit is composed of four domains (Fig. 7.11):

- an N-terminal domain
- two AAA domains called AAA-1 and AAA-2, oriented in a head-to-tail arrangement
- a linker called middle (M) domain.

The crystal structure revealed that the ClpB hexamer has a two-tiered ring structure, with a diameter of 140–140 Å and a height of 90 Å, where AAA-1 is located directly above AAA-2 of the same subunit in the hexameric model. The two AAA domains are connected via a short helix which may act as a hinge, allowing large movements of the AAA domains during the nucleotide cycle. The N domain is connected via a flexible linker to AAA-1 and is very mobile with a large rotational freedom. The function of the N domain remains elusive. The M domain is inserted at the end of the AAA-1 domain and forms a large coiled-coil domain. It is located on the outer surface of the ClpB molecule and is proposed to act as a molecular crowbar, enabling ClpB to convert large protein aggregates into smaller ones. The diameter of the central pore (16 Å) is sufficient to allow translocation of a peptide stretch of a substrate peptide into the central cavity. But ClpB by itself does not act on protein aggregates: it needs the DnaK chaperone team.

Based on the available information, a mechanistic model has been proposed in which large protein aggregates can be dissolved by the concerted action of ClpB and the DnaK system (Fig. 7.12). ClpB binds to large protein aggregates and dissolves them into smaller aggregates, mediated by movements of the M domain acting as a crowbar. Furthermore, ClpB can bind to these medium-sized aggregates to extract unfolded proteins and transfer them to the DnaK team. Unfolded polypetides or small-sized aggregates serve as substrates for the DnaK system, ensuring substrate refolding into their native three-dimensional structure.

Fig. 7.11 Domain organization of the ClpB protein. The ClpB protein is composed of four domains: an N-terminal domain of unknown function fused to the AAA-1 domain through a flexible linker. The M domain is inserted at the end of the AAA-1 domain, forms a large coiled-coil structure and is located on the outer surface of the molecule. A second AAA domain, the AAA-2 domain is located at the C-terminal end of ClpB. A. Mogk, B. Bukau **2004**, *Curr. Biol.* 14, R78–R80; Fig. 1A.

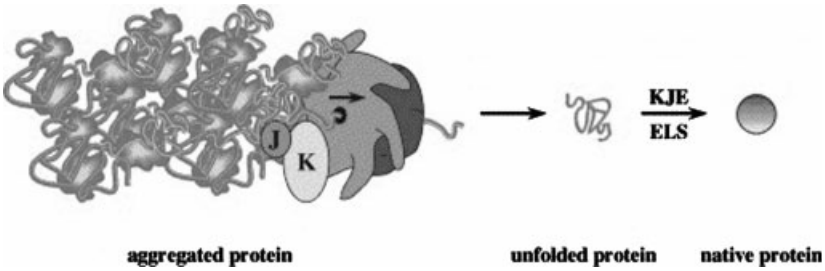


Fig. 7.12 Hypothetical sequential model for protein disaggregatin by the ClpB-DnaK complex. The ClpB-DnaK complex recognizes large protein aggregates, and the ClpB hexamer pulls apart the aggregate during the ATPase cycle of ClpB. Next, the DnaK or the GroE team promote refolding of single unfolded polypeptide chains. J. Weibezahn et al. **2004**, *Cell* 119, 653; Fig. 6.

7.2.4 Intramolecular Chaperones

A number of proteins, mainly extracellular proteases, contain their own chaperones. These proteins are synthesized as preproproteins in the cytoplasm and translocated through the cytoplasmic membrane using the Sec pathway (see Section 8.4.3). During the translocation process, the presequence is cleaved off by the signal peptidase. The prosequence at the N-terminus is essential for the correct folding of the remainder of the chain, but is then cleaved off and therefore absent from the final active protein. Such prosequences are found during the synthesis of extracellular serine proteases, such as subtilisin E and α-lytic protease, and have been termed *intramolecular chaperones*. If the final protein is produced without its prosequence, it is unable to fold correctly and gain activity. If purified prosequence is added, it will be able to fold into its native three-dimensional structure. One protease, the maturation of which has been studied in detail, is subtilisin E.

Subtilisin E is an alkaline serine protease found in *B. subtilis*. In the cytoplasm, this protein exists as a precursor termed preprosubtilisin. It consists of a 29-residue signal peptide, a 77-residue propeptide and a 275-residue protease domain. The presequence acts as a signal peptide and facilitates the secretion of prosubtilisin across the cytoplasmic membrane, whereas the proregion or propeptide functions as an intramolecular chaperone that guides correct folding of the subtilisin domain. Upon completion of folding, the covalent link between the mature domain and the intramolecular chaperone is hydrolyzed via autoprocessing, a step accompanied by substantial structural reorganization. After autoprocessing, the cleaved 77-residue intramolecular chaperone remains bound to the active site of subtilisin. Active mature subtilisin is therefore obtained through autoproteolytic degradation of the cleaved intramolecular chaperone domain. In the absence of the intramolecular chaperone domain, the subtilisin domain is trapped in a stable, but inactive, molten globule-like intermediate state. Stoichiometric addition of the intramolecular chaperone domain to this intermediate *in trans* facili-

tates its conversion into a proteolytically active state. Although initially identified in serine proteases, such as subtilisin, aqualysin I (thermostable subtilisin homolog secreted by *T. aquaticus*) and α-lytic protease, intramolecular chaperone-mediated protein folding is now shown to exist in several other protease families. These families include cysteine proteases, aspartic proteases and metalloproteases. Besides playing a crucial role during folding, the prosequence exhibits another important function. As long as it is covalently attached to the mature polypeptide, it inhibits its proteolytic function, thereby rendering it inactive in the cytoplasm. Furthermore, the prosequence may assist in secretion of the protein.

J. Eder, A.R. Fersht **1995**, Pro-sequence-assisted protein folding, *Mol. Microbiol.* 16, 609–614.

7.2.5
Folding of Nascent Polypeptide Chains

An average *E. coli* cell growing logarithmically at 30–37 °C on a glucose minimal medium contains approximately 2.35×10^6 polypeptide chains, with an average size of ~35 kDa (317 residues). The total number of different polypeptide chains found in the cytoplasm is ~2600. Besides these millions of polypeptide chains, these cells contain 2–4 copies of the chromosome, about 20000 ribosomes, tRNA and small RNA molecules and metabolites, creating a highly crowded environment and leading to *molecular crowding*. How can polypeptide chains fold correctly in such a crowded environment?

Protein Folding in the Cytoplasm

Folding of newly synthesized cytosolic proteins is principally orchestrated by the three major chaperones, namely trigger factor (TF), DnaK/DnaJ/GrpE and GroEL/GroES. The majority of the nascent polypeptides interact cotranslationally with the ATP-independent ribosome-associated TF and reach their native state without further folding assistance (Fig. 7.13). A substantial number of nascent polypeptides need the help of either the DnaK or/and the GroE chaperone system. While the DnaK chaperone machine interacts in a co- or post-translational mode, with nearly 15% of newly synthesized proteins having a molecular mass of >30 kDa, the GroE system interacts post-translationally with about 10% of the newly synthesized polypeptides (20–60 kDa). TF and the DnaK system often cooperate in the folding of newly synthesized proteins, as outlined above. In the absence of the TF, the number of newly synthesized polypeptides interacting with DnaK is greatly enhanced from 15% to about 40%. Deletion of both *dnaK* and *tig* in *E. coli* causes synthetic lethality, but cells can tolerate deletion of both genes at temperatures <30 °C. Under these conditions, they accumulate high levels of aggregated proteins at the permissive temperatures, suggesting that other chaperones might be involved in protein folding. It can be shown that the secretion-dedicated chaperone SecB efficiently suppresses both the temperature sensitivity and the aggregation-prone phenotypes of strains devoid of *dnaK* and *tig*.

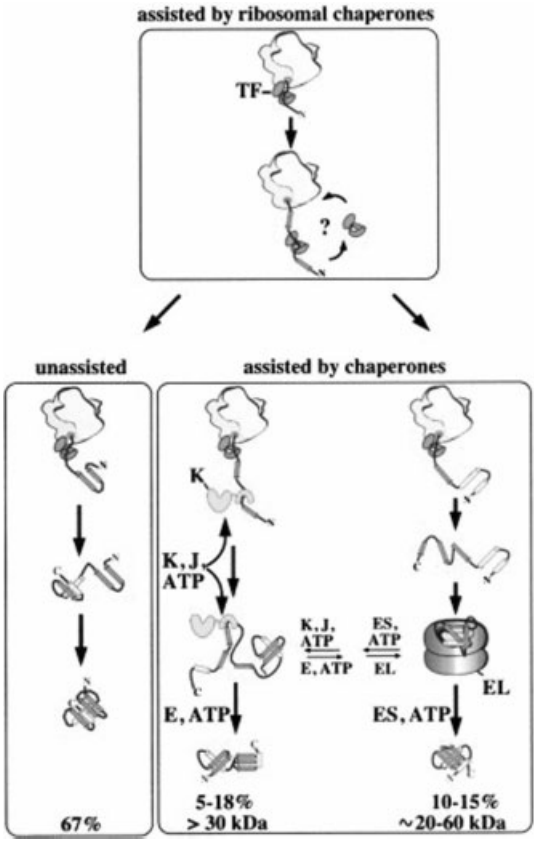


Fig. 7.13 Model for chaperone-assisted folding of nascent polypeptide chains in the cytoplasm. Nascent polypeptide chains first interact with the trigger factor (TF) loacted at the exit tunnel of the ribosome. While most proteins are able to fold in the complete absence of any folder chaperone, 5–18% are dependent on the DnaK and another 10–15% on the GroE team with overlapping specificity. B. Bukau, et al. **2000**, *Cell* 101, 119–122; Fig. 1. (This figure also appears with the color plates.)

Protein Folding in the Periplasm

Proteins with an ultimate destination in the periplasm and outer membrane are synthesized as precursors with a cleavable N-terminal signal sequence. Translocation of most proteins occurs by the Sec pathway across a channel formed by the SecYEG complex, while translocation of some containing two consecutive arginine residues near the N-terminus of their signal sequence utilize the alternative Tat pathway (see Chapter 8). The folding of proteins residing in the periplasm of Gram-negative bacteria is less well understood. None of the folder chaperones has been identified in the periplasm, as they are ATP-dependent and ATP is absent from this compartment. But folding catalysts have been identified in the periplasm, contributing rather differently from the chaperones in the protein-folding reaction. One class of protein-folding catalysts are the *p*rotein *d*isulfide *i*somerases

(PDIs) which perform thiol–disulfide exchanges. The second class consists of peptidyl prolyl isomerases (PPIases) which isomerize Xaa-Pro peptidyl bonds (see below for a detailed description of these two classes of protein-folding catalysts).

A third class of periplasmic proteins act as chaperones, such as the *E. coli* SurA and FkpA, which are involved in the folding and assembly of outer membrane proteins. In addition, *in vitro* chaperone activity of several periplasmic proteins such as DsbC and DsbG or substrate-binding proteins has been reported, but their *in vivo* contribution is not yet clear. The SurA protein consists of four separate domains following the signal peptide: an N-terminal domain of ~150 amino acids, followed by two peptidyl–prolyl isomerase domains of ~100 residues each (PPIase P1, P2) and finally, a C-terminal domain of ~40 amino acids. SurA has affinity for peptides with a consensus sequence of the form aromatic/polar/aromatic/nonpolar/proline, a peptide motif that is characteristic of integral outer membrane proteins. In a *surA* mutant, maturation of outer membrane proteins is impaired, and OmpA, OmpC and LamB are unable to achieve complete folding. If SurA is present at an enhanced copy number, it can promote the folding of many otherwise unstable proteins. FkpA is a homodimer with subunits divided into two domains where the N-domains form the dimeric interfaces. The cleft formed by association of the N-domain has an intermediate negative potential and is assumed to accept unfolded polypeptide substrates. An alternative fate for non-native periplasmic proteins is degradation by one of the ten proteases identified so far in *E. coli* (see below).

M. Miot, J.-M. Betton **2004**, Protein quality control in the bacterial periplasm, *Microbiol. Cell Fact.* 3, 4.

7.3 Protein-folding Catalysts

The folding of most proteins translocated into the periplasm encounters two types of rate-limiting steps, which are overcome by two classes of catalysts: protein disulfide isomerases (PDI) and peptidyl-prolyl *cis-trans* isomerases (PPIase). In *E. coli*, there are at least six known Dsb poteins that are involved in the oxidation of disulfide bonds or the rearrangement of wrongly paired disulfides. PPIases are found in both the cytoplasm and the periplasm and catalyze the rapid interconversion between the *cis* and *trans* forms of the peptide bond Xaa-Pro.

7.3.1 Peptidyl-prolyl Isomerases

PPIases were found to belong to three families, termed cyclophilins (Cyp), FK506-binding proteins (FKBPs) and parvulins. The FKBPs are a family of PPIases that have been named due to their ability to bind the inhibitor FKB506, a macrolide isolated from *S. tsukubaensis*. This inhibitor was initially discovered for its immunosuppressive effect, and since then several analogs have been isolated from different

Table 7.2 The PPIases of *E. coli*.

Name	*Location*	*Family*	*Remarks*
FkpA	Periplasm	FKBP	Member of the sigma-E regulon
PpiA (RotA)	Periplasm	Cyclophilin	Member of the Cpx regulon
SurA	Periplasm	Parvulin	Correct folding of OMP monomers
PpiD	Periplasm	Parvulin	
PpiC	Cytoplasm	Parvulin	
Tig (TF)	Cytoplasm		Trigger factor; ribosome-associated
FklB	Cytoplasm	FKBP	
SlyD	Cytoplasm		

Streptomyces strains, such as FK520 from *S. hygroscopius*. Because of their inhibition by immunosuppressant drugs (cyclosporin A or FKB506), the first two families are also called immunophilins. Despite their ubiquitous occurrence, however, most of these enzymes are not essential under normal growth conditions. *In vitro*, the PPIases have been initially characterized by their ability to catalyze peptide bond rotation in a short tetrapeptide model substrate to become sensitive to chymotrypsin. The PPIases found in *E. coli* are summarized in Table 7.2.

Although all PPIases catalyze the rate-limiting step in the refolding of RNase T1 *in vitro*, their cellular role remains enigmatic. With the exception of *surA*, the absence of a significant phenotype for null mutants indicates that these periplasmic PPIases are not essential for viability or have overlapping functions.

7.4 Disulfide Bond Formation

A key step in protein folding is the formation of disulfide bonds between cysteine residues, stabilizing their tertiary structure. Disulfide bond oxidation, reduction and isomerization are catalyzed processes, facilitated by members of the thioredoxin superfamily. Though disulfide bonds are essential in stabilizing polypeptides, they are rarely found in cytosolic proteins. Instead, disulfide bonds usually occur in extracytoplasmic proteins. Why do disulfide bonds not occur in cytoplasmic proteins?

- A number of cytoplasmic enzymes (ribonuclease reductase, methionine sulfoxide reductase) rely on reduced cysteines in their active site. The presence of oxidized disulfides could lead to inactivation of these enzymes.
- A partially unfolded conformation is required for the translocation of most proteins through the inner membrane. Disulfide bonds would hinder the translocation step.
- A number of virulence factors and toxins contain multiple disulfide bonds. Therefore, prevention of disulfide bond formation within the cytoplasm renders these proteins inactive.

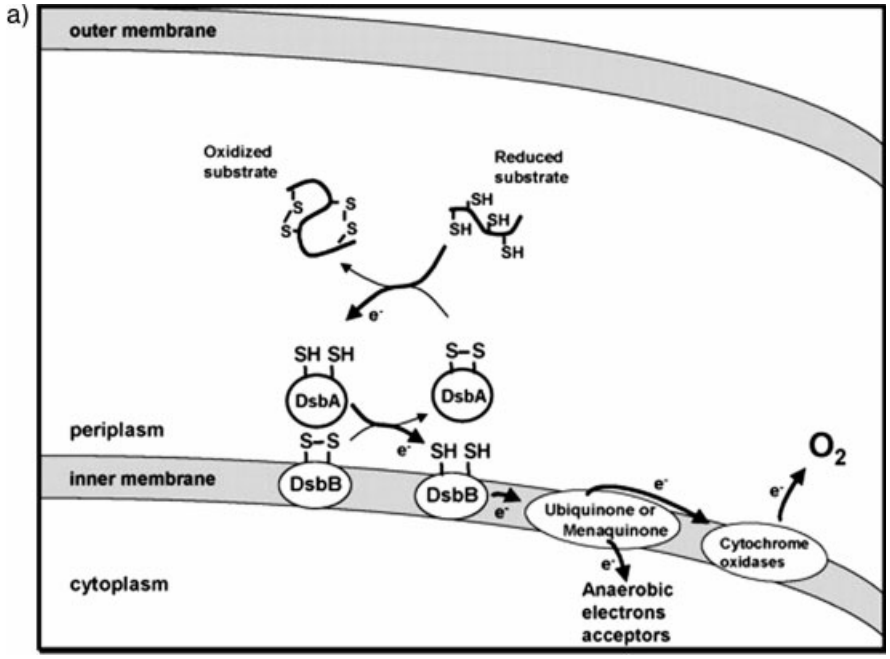
To prevent disulfide bond formations within the cytoplasm, two impediments have evolved. One important barrier is the extremely reducing environment of the cytoplasm, maintained by the thioredoxin/thioredoxin reductase, glutathione/glutathione reductase and glutaredoxin/glutaredoxin reductase systems, ensuring that cytoplasmic cysteine residues are kept in a reduced state. The second barrier is the absence of enzymes catalyzing disulfide bond formation, but these enzymes are present in the periplasm of Gram-negative bacteria. Proteins that contain disulfide bonds can be divided into two classes:

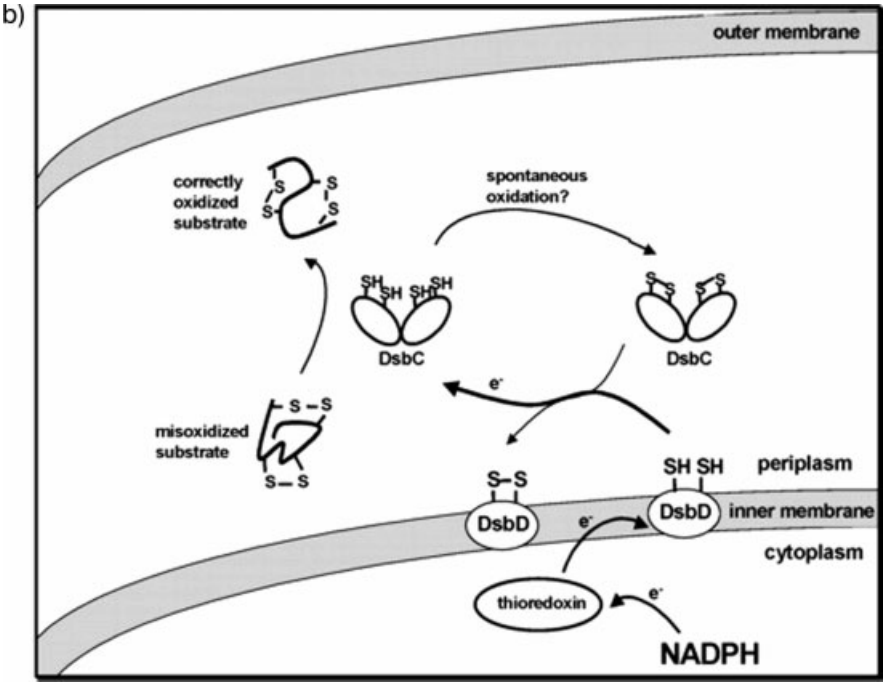
1. For class I proteins, the cysteine–cysteine is a stable part of their final folded structure; and the disulfide bond may contribute to the folding pathway of the protein and to the stability of its native state.
2. For class II proteins, pairs of cysteines alternate between the reduced and oxidized states; and the oxidative–reductive cycling of the disulfide bond may be central to a protein's activity as an enzyme (e.g., certain ribonucleotide reductases) or may be involved in a protein's activation and deactivation (e.g., OxyR).

Proteins that are capable of catalyzing protein disulfide bond formation are called thiol-disulfide oxidoreductases or *p*rotein *d*isulfide *i*somerases (PDIs). Many of them belong to the thioredoxin superfamily, which is defined by an active site containing a CXXC motif and by a thioredoxin fold. While these proteins localized in the cytoplasm perform mainly reductive steps, those in extracytoplasmic compartments can act as oxidants. Additional enzymes not being members of the thioredoxin family use redox active cysteine pairs that are separated by more than two amino acids. In addition, these enzymes may use small-molecule electron donor and receptor cofactors, such as FAD, NADPH, NADH, quinones and lipoic acid.

In the *E. coli* periplasm, disulfide bond formation and isomerization are catalyzed by two separate pathways. While the DsbA-DbsB (*d*isulfide *b*ond) pathway oxidizes thiol groups to form disulfide bonds *de novo*, the DsbC-DsbD pathway isomerizes mismatched disulfides. DsbA, a monomeric 21-kDa protein present in the periplasm, is the primary oxidant of proteins, interacting with substrate proteins and oxidizing them. DsbA is member of the thioredoxin family and contains two active-site cysteine residues in the CXXC motif embedded in the thioredoxin-like fold. These two cysteines are found in the disulfide-bonded form *in vivo*. DsbA rapidly oxidizes proteins translocated into the periplasm by donating its disulfide to pairs of cysteines in substrate proteins by a thiol-disulfide exchange reaction (Fig. 7.14a).

Fig. 7.14 (a) The DsbA–DsbB pathway catalyzes the formation of disulfide bonds. DsbA interacts with the target protein containing reduced cysteines and oxidizes them to a disulfide bond. DsbA is subsequently reduced by transferring its electrons to DsbB. Under aerobic conditions, DsbB donates the electrons to ubiquinone which transfers them to cytochrome oxidases and finally of molecular oxygen. Under anaerobic conditions, DsbB donates the electrons to menaquinone which passes them over to anaerobic electron acceptors such as fumarate or nitrate reductase. A. Hiniker **2003**, *Biochemistry* 42, 1179–1185; Fig. 1. (b) Disulfide bond formation in the periplasm of *E. coli*. DsbC recognizes target proteins with non-native disulfide bonds, allowing rearrangement of these bonds into their native pairings. DsbC is maintained in its reduced state by DsbD which receives the electrons from NADPH via thioredoxin. A. Hiniker **2003**, *Biochemistry* 42, 1179–1185; Fig. 2.


a)
outer membrane
Oxidized substrate
Reduced substrate
SH SH
S–S
DsbA
periplasm
inner membrane
DsbB
Ubiquinone or Menaquinone
Cytochrome oxidases
O_2
cytoplasm
Anaerobic electrons acceptors
e-


b)
outer membrane
correctly oxidized substrate
spontaneous oxidation?
DsbC
misoxidized substrate
periplasm
inner membrane
cytoplasm
DsbD
thioredoxin
NADPH
e-

Oxidized DsbA reacts with translocated proteins to form mixed disulfides, then releases substrates in the oxidized form. DsbA recognizes a wide variety of proteins localized to the bacterial cell envelope. This is accomplished because the oxidized form is more flexible than the reduced one; and the rigid nature of the reduced form might facilitate the release of the oxidized products. As the result of the transfer of the disulfide bond of DsbA to a substrate protein, the two cysteines in the active site of DsbA become reduced and have to become reoxidized. This oxidation step is carried out by a cytoplasmic membrane protein, DsbB, which restores the disulfide bond to DsbA by using the oxidizing power of the electron transport chain. Under aerobic conditions, DsbB transfers two electrons from DsbA onto oxidized ubiquinone. The reduced ubiquinone is then reoxidized by the terminal cytochrome oxidases *bd* and *bo*, which finally transfer the electrons to oxygen. Under anaerobic conditions, DsbB passes the electrons from DsbA to menaquinone and from there to final electron acceptors other than oxygen. DsbB (20 kDa) spans the cytoplasmic membrane four times, with both N and C termini facing the cytoplasm. Each periplasmic domain contains one pair of essential cysteines which are disulfide-bonded *in vivo* and redox active. This enzyme becomes reduced after it reoxidizes DsbA. How is DsbB reoxidized? Under aerobic conditions, electrons are transferred from DsbB to ubiquinine, then to cytochrome oxidases and finally to molecular oxygen. Under anaerobic conditions, reoxidation occurs by electron flow from DsbB to menoquinone to terminal acceptors such as fumarate or nitrate reductase.

As the number of cysteine residues in a protein increases, the number of possible disulfide bonds increases, too, including wrong pairings. How does the cell catalyze formation of the correct cysteine pairings? It is assumed that DsbA introduces disulfide bonds into the partially unfolded polypeptides during translocation into the periplasm, suggesting that DsbA may introduce disulfides between sequential cysteine residues, even if these do not exist in the native protein. Correct disulfide bond formations are controlled by the DsbC-DsbD pathway. DsbC is a V-shaped homodimer with each monomer (23.4 kDa) forming one arm of the V. Each DsbC monomer consists of two separate domains, an N-terminal dimerization domain joined by a linker helix to a thioredoxin-like C-terminal domain containing the CXXC active site; and the two active sites of the dimer function independently of each other. The dimeric arrangement of the molecule is fundamental for its proposed roles in the periplasm as a thiol-disulfide bond isomerase and as a chaperone. Incorrect disulfide bonds in proteins, such as RNase A and urokinase can be successfully corrected by DsbC. As a chaperone, DsbC assists in the reactivation *in vitro* of denatured proteins that do not contain cysteine residues. This dual role presumably allows DsbC synergistally to identify (chaperone activity) and unfold non-native disulfide-bonded intermediates (isomerase activity). The recognition of non-native proteins probaly resides in the uncharged cleft embodied by the two arms of the V-shaped molecule. This cleft can switch between open and closed conformations, depending on the redox state of their active thioredoxin folds. This flexibility allows DsbC to adjust the cleft to the size of the target protein. After binding of the substrate protein with incorrect disulfide bonds, DsbC performs a nucleophilic attack of the N-terminal cysteine of the

CXXC active site onto the substrate protein (Fig. 7.14b), forming a mixed disulfide between DsbC and substrate. Attack of a third substrate thiol group onto the mixed disulfide allows resolution of the mixed disulfide and thereby rearrangement of the substrate disulfides to a more stable conformation. The *E. coli* periplasm contains a second disulfide isomerase called DsbG; and the major difference between DsbC and DsbD seems to be just their substrate specificity. Both DsbC and DsbG are maintained in their reduced form by DsbD, another inner membrane protein. The current model for maintaining DsbC and DsbG in the reduced form involves the passage of electrons from the reducing environment of the cytoplasm to the oxidizing periplasm. Electrons are passed from thioredoxin kept in reduced form by thioredoxin reductase and NADPH to DsbD. The 59-kDa DsbD protein consists of three domains, an N-terminal periplasmic domain, a hydrophobic transmembrane domain with eight transmembrane segments and a C-terminal thioredoxin-like domain, also located in the periplasm. Each domain has a pair of conserved cysteine residues that participate in consecutive disulfide exchange reactions to transfer electrons from the cytosol to the periplasm. The electrons are transferred from thioredoxin to the transmembrane domain, next to the C-terminal domain and finally to the N-terminal domain which reduces DsbC or DsbG.

H. Kadokura, et al. **2003**, Protein disulfide bond formation in prokaryotes, *Annu. Rev. Biochem.* 72, 111–135.

7.5 Proteases

Proteases in general refer to a group of enzymes whose catalytic function is to hydrolyze peptide bonds, leading to the breakdown of their substrates. They can be divided into five different groups, depending on the active site residue or ion that carries out hydrolysis of the peptide bonds: (a) serine, (b) threonine, (c) cysteine, (d) aspartic and (e) metalloproteases. The physiological roles of proteases are very diverse, ranging from digestive functions, protein maturation, the precise processing of regulatory proteins and removal of damaged proteins. These damaged or non-native proteins can represent a serious problem to each cell, as they might accumulate as large aggregates, a process associated with prion and other amyloid diseases, leading to cell death. *In vivo*, the formation of protein aggregates is favored due to the extremely high intracellular protein concentrations (100–150 mg ml–1). To prevent the formation of aggregates, cells have developed an elaborate protein quality control system where molecular chaperones and proteases act together to reduce the amount of denatured proteins within the cells. These quality control systems can distinguish the substrates that can be refolded from the severely damaged proteins ready for degradation. Proteases do not attack their protein substrate at random, but they exhibit a high degree of specificity in identifying and thereby binding their substrates. Substrate specificity is often defined by the structural properties of the active site or sometimes by *adaptor pro-*

teins that bind and feed substrates to the protease. Recognition sites can be located either at the N- or the C-terminal end of a substrate protein. With some proteins, recognition sites may be located in the central part of the protein (*cryptic cleavage sites*) and become accessible as a result of unfolding, loss of an interaction partner or cleavage by another protease. It has been shown that the N-terminal amino acid residues can strongly influence the halflife of a model protein. While N-terminal arginine, lysine, leucine, phenylalanine, tyrosine and tryptophan confer halflives of 2 min, the remaining amino-terminal residues confer halflives greater than 10 h, on the same model protein. This observation is termed the *N-end rule*; and it operates in all eukaryotes examined. The N-end rule relates the metabolic stability of a protein to the identity of its N-terminal residue, named *N-degron*. Many of the classic proteases such as trypsin are synthesized in an inactive form called zymogen that are converted into the active form by proteolytic cleavage of a propeptide. A new principle is emerging, indicating reversible zymogen activation where either other proteins or external factors (such as concentrated salts or temperature) can switch protease activity on or off.

As to their dependence on energy, proteases can be divided into two groups, namely ATP-dependent and ATP-independent proteases. While ATP-dependent proteases exclusively reside in the cytoplasm where ATP is synthesized, ATP-independent proteases are found outside of the cytoplasm either in the periplasm of Gram-negative bacteria or anchored in the inner or outer membrane or even secreted into the medium. In the following sections, I will first deal in detail with the ATP-dependent proteases and then mention some important energy-independent proteases.

7.5.1 ATP-dependent Proteases

ATP-dependent proteases play pivotal roles in the regulation of various cellular events, including the removal of non-native proteins, cell division, signal transduction and stress responses, by rapidly changing the levels of key regulatory proteins. Eubacterial cells are equipped with several ATP-dependent proteases. In *E. coli*, five enzymes catalyze ATP-dependent proteolysis: Lon, ClpAP, ClpXP, ClpYQ (HslUV) and FtsH, which are classified as one- and two-component proteases. Lon and FtsH are one-component proteases where a single polypeptide chain carries both ATPase and proteolytic sites. In contrast, the ATPase and the proteolytic domain are present on separate subunits in the two-component proteases. In *E. coli*, only FtsH is essential, while ClpXP or Lon are essential in other species. Identifying the proteolytic targets of specific proteases is critical to any general understanding of their diverse cellular functions and to provide a way to decipher the rules by which these enzymes recognize substrates. In addition, proteolysis plays a critical role in many regulatory circuits, keeping basal levels of regulatory proteins low and rapidly removing proteins when they are no longer needed. These proteases form molecular cages by association of their proteolytic subunits into oligomeric rings. The proteolytic sites are located on the inner wall of the cen-

tral cavity and are accessible through narrow axial lateral pores. By this measure, folded proteins are excluded from the lumen of the complex.

The current model for degradation by ATP-dependent proteases involves three steps:

1. Recognition: the protease first selects a protein for degradation, either because it has an accessible tag that specifies protease binding, or because an internal degradation signal has become exposed. Recognition is the main commitment step.
2. Translocation: following recognition, ATP hydrolysis promotes both unfolding of the substrate and its subsequent translocation into the proteolytic chamber (dual role of ATP).
3. Proteolysis: following translocation, the protein is hydrolyzed to small fragments, which are released from the chamber into the cytoplasm.

ATPγS supports the assembly of complexes and the binding of substrate, but not translocation and degradation (or very slowly). Small peptides easily diffuse into the proteolytic chamber and are degraded in the presence of ATPγS.

The Lon Protease

Lon, the first of the ATP-dependent proteases to be studied, is a soluble cytoplasmic serine protease. Genes encoding Lon protease have been found in a large variety of prokaryotes (some species such as *B. subtilis* contain two *lon* genes, named *lonA* and *lonB*), where they have been implicated in developmental pathways as well as in the degradation of specific regulatory proteins. The 94-kDa *E. coli* polypeptide chain consists of a coiled-coil domain at its N-terminus, followed by the Walker A and B motifs involved in ATP binding and hydrolysis and the peptidase domain with the serine active site (S679) forming the C-terminal domain (Fig. 7.15). Although the three-dimensional structure of Lon is not yet available, preliminary data suggest that Lon assembles into a ring-shaped hexamer. Lon seems to use its ATPase domain to unfold substrates and then translocate them to the proteolytic domain. Consistent with this hypothesis, the 72-amino-acid CcdA substrate is degraded in the presence of hydrolyzable ATP, while a shorter, less-structured derivative or partially denatured CcdA is degraded independent of ATP hydrolysis.

Lon is involved in the control of many different cellular processes, such as protein quality control, UV tolerance, capsule synthesis, phage development, heme

Fig. 7.15 Structural features of the Lon protease. The 94-kDa enzyme (784 residues) is composed of a coiled-coil structure at the N-terminus, followed by a Walker A and B box in its central part and the peptidase domain with the serine active site (S679). D.A. Dougan, et al. **2002**, *FEBS Lett.* 529, 6–10; Fig. 1.

and methionine biosynthesis, cell cycle and differentiation, and ribosomal protein degradation after amino acid starvation, among other processes. Mutations in *lon* reduce the turnover of about 50% of the abnormal proteins in *E. coli* significantly, suggesting an important role in general protein quality control. In addition to the substrates already mentioned, Lon is responsible for degradation of several regulatory proteins, such as SulA, RscA and λN. SulA is one of the most physiologically important substrates. The *sulA* gene is transcriptionally induced by environmental stresses, such as UV irradiation, and prevents premature segregation of damaged DNA into daughter cells during DNA repair processes. Induced SulA prevents the self-assembly of the FtsZ protein (see Section 5.2.9), leading to the inhibition of cell division. Lon preferentially degrades certain sites of SulA with an apparent consensus sequence, Leu-Ser, in the central and C-terminal regions which are important for the function of SulA, in the presence of ATP. Among them, six peptide bonds are cleaved preferentially in the early stage of digestion. UmuD, another substrate, is recognized by virtue of a transferable motif located within the first 24 amino acids, whereas two other Lon substrates, CcdA and SulA, have their motif present in the C-terminal regions.

Overproduction of Lon is lethal for bacteria, suggesting that one or more essential protein(s) is (are) substrate(s) of Lon. Lon overproduction specifically inhibits translation through at least two different pathways. One of these pathways is the chromosomal *yefM-yoeB* toxin–antitoxin system, where the YoeB toxin induces cleavage of translated mRNAs (see Section 3.4.2). Lon overproduction specifically activates YoeB-dependent mRNAs cleavage. Based on this observation, it has been suggested that translation inhibition (mediated by Lon degradation of a component of the YoeB-independent pathway) in turn activates the YoeB toxin.

One inhibitor of the Lon protease has been identified encoded by the *pinA* (for *p*roteolysis *in*hibition *A*) gene. PinA is a heat-stable protein of 18.8 kDa and binds with high affinity to Lon in a 1:1 stoichiometry (one PinA per one Lon homohexamer). PinA inhibits reversibly the protein-activated ATPase activity at 80–95%, by interacting at a novel regulatory or enzymatic site involved in coupling between ATP hydrolysis and proteolysis.

The Clp Proteases

Three different two-component Clp proteases have been described in *E. coli*: ClpAP, ClpXP and ClpYQ (HslUV). The three ATPase components ClpA, ClpX and ClpY form ring-like structures composed of six subunits. When complexed with their proteolytic component, ClpP or ClpQ, the ATPase rings are positioned at one or both ends of the proteolytic component and regulate degradation by binding, unfolding and translocation of specific substrates into the proteolytic chamber. The endopeptidase active sites of both protease components are located in an internal cavity generated by two stacked rings of six or seven identical subunits. The substrate is degraded into peptides which diffuse out. There is a symmetry mismatch in the ClpAP/ClpXP systems, with ClpA and ClpX forming a hexamer and ClpP forming a heptamer. No such symmetry mismatch occurs in

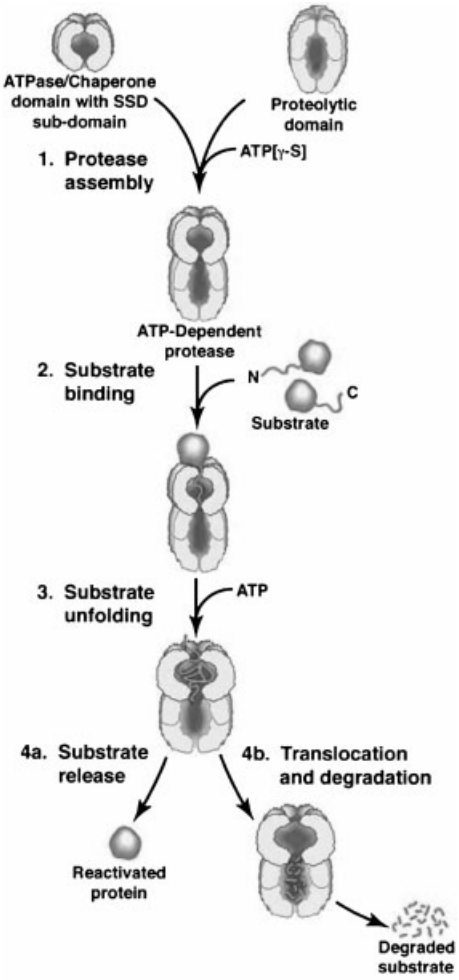


Fig. 7.16 Model of substrate recognition by Clp proteases. 1. Protease assembly. In a first step, association between one chaperone ring and two proteolytic rings occurs. This step does not need ATP hydrolysis since it can occur in the presence of nonhydrolyzable ATPγS. 2. Substrate binding. Interaction between substrate proteins and the chaperone occurs at the *s*ensor and substrate *d*iscrimination (SSD) domain. 3. Substrate unfolding. Conformational changes within substrate proteins need ATP hydrolysis. Then, substrate proteins are either remodeled and released from the chaperone (4a) or translocated into the proteolytic chamber and degraded into peptides. S. Wickner **1999**, *Proc. Natl. Acad. Sci. USA* 96, 8318; Fig. 1.

ClpQY where both ClpQ and ClpY are hexamers. These complexes couple the protein unfolding activity of their molecular chaperone partner with the degradative activity of an endopeptidase.The current view of substrate recognition, unfolding and translocation into the proteolytic chamber, followed by degradation into peptides is shown in Fig. 7.16.

The ATPase components recognize a *recognition motif* which can be located either at the N- or the C-terminus of the substrate protein and which constitutes a *transferable signal*. For the MuA protein, a substrate for the ClpXP protease, the recognition motif is located within the C-terminal ten amino acids. In the case of the phage λO protein, targeting to the ClpXP occurs via sequences contained within the N-terminus. In the case of the P1 RepA protein specifically recognized by ClpAP, the recognition sequence is located within the N-terminal 20 amino acids.

The ClpAP Protease

The ClpA ATPase component can associate with the proteolytic ClpP component which is a tetradecamer arranged into two rings of seven subunits stacked on the axis of symmetry, where the proteolytic component is flanked by one ATPase component at either or both ends. ClpP contains many active sites involved in cleaving peptide bonds which are buried on the inside of the cylindrical chamber. In the ClpAP complex, ClpA recognizes a substrate protein and translocates it into the ClpP chamber for degradation. Since the pores at the ends of the ClpP ring are not wide enough to admit a globular folded protein, ClpA has to unfold the substrate protein and translocate the elongated polypeptide chain through the pore into the proteolytic chamber.

The ClpXP Protease

E. coli ClpXP is an ATP-dependent intracellular protease composed of an ATP-dependent protein unfoldase, the ClpX hexamer and a double-ring serine protease, ClpP, where each ring contains seven subunits. ClpX selects substrates for degradation, unfolds them and then translocates the unfolded polypeptide into a chamber within ClpP, where degradation occurs. ClpXP degrades different substrate proteins at different times, depending on growth and environmental conditions. One fundamental step during substrate recognition involves the binding of a substrate-processing site on ClpX to a peptide degradation signal, which is often located at or near the N- or C-terminus of the target protein. Peptide degradation sequences may be constitutively recognized or become accessible to ClpX only after cleavage by another protein or after a conformational change. After recognition of the peptide degradation signal by the ClpX processing site, ATP-dependent conformational changes in ClpX are thought to generate a transient pulling force that destabilizes the attached native protein. By using repeated cycles of ATP hydrolysis, ClpX unfolds the protein substrate and translocates it into ClpP for degradation. While some peptide degradation signals are sufficient to cause virtually any attached polypeptide to become efficiently proteolyzed by ClpXP (proteins with an SsrA-tag; see below), a second mode of substrate recognition by ClpX involves binding sites that interact with so-called *adaptor proteins*. These adaptor proteins enhance degradation of specific substrates without being degraded (see below).

The following ClpXP-specific substrates have been identified: the phage λO replication protein, phage Mu repressor and MuA transposase, RK2 replication protein TrfA, P1 antidote protein PhD, stationary phase sigma factor σ^S and the UmuD′ subunit of error-prone DNA polymerase. Furthermore, ClpXP is involved in the removal of LexA repressor protein fragments after RecA processing and the type I restriction-modification subunit HsdR.

The ClpYQ (HslUV) Protease

The 50-kDa ClpQ (HslU) ATPase forms a hexameric ring bound to either one or both ClpY hexameric rings. The 19-kDa ClpY (HslV) peptidase is homologous to

certain β-type subunits of the 20S eukaryotic proteasome. The *clpQY* genes are induced by heat shock, implying that the enzyme is involved in the degradation of damaged polypeptides. The enzyme degrades several short-lived regulatory proteins including SulA, but ClpQY seems to act as a backup system for Lon. During breakdown of SulA, ClpQY produces 58 peptides with various sizes (3–31 residues), where cleavage occurred at 39 peptide bonds, preferentially after Leu.

The conserved tyrosine 91 residues that line the central pore of the ClpY hexamer are crucial for substrate degradation and exhibit an apparent conformational flexibility. As the positioning of such aromatic residues at the central pore is conserved in many AAA^+ proteins, it has been suggested that they act as molecular clamps by binding and releasing substrates in a nucleotide-dependent manner. Such a mechanism would directly couple substrate translocation to substrate unfolding.

The FtsH Protease

FtsH (HflB) is the only membrane-anchored ATP-dependent protease with its active site facing the cytoplasm. The 71-kDa protein contains two transmembrane segments near its N-terminus followed by a well conserved 200-aa AAA motif with the Walker A and B boxes and Zn^{2+}-binding motif which is part of the metalloprotease domain (Fig. 7.17). FtsH forms a hexameric ring-like structure that is crucial for its ATPase and proteolytic activities. The N-terminal region exerts dual functions in membrane anchoring and homooligomerization. In *E. coli*, FtsH forms a complex with the membrane complex HflKC, supposed to have a regulatory role in the proteolytic function of FtsH. Interaction between FtsH and the HflKC complex occurs between the short extracytoplasmic loop of FtsH and the periplasmically exposed HflKC complex.

FtsH is the only protease completely conserved among all eubacteria examined to date and is essential for viability in *E. coli*. Homologs of FtsH are present in both mitochondria and chloroplasts. FtsH, like other ATP-dependent proteases, forms a hexameric ring structure with its proteolytic active site buried within the central cavity. Since this proteolytic chamber is accessible only to unfolded and extended polypeptides, a topological barrier between cytoplasmic proteins and the site of proteolysis is created, thereby preventing aberrant degradation of cellular proteins. FtsH degrades both integral membrane and cytoplasmic proteins. The major housekeeping function of FtsH is to sequentially degrade polytopic membrane proteins such as the SecY subunit of the protein translocase (see Section 8.4.1) and subunit *a* of the proton ATPase F_0F_1 when they are not associated with

Fig. 7.17 Structural features of the FtsH metalloprotease. Two transmembrane segments at the N-terminus are followed by a AAA domain with the Walker A and B motifs and a zinc-binding motif within the C-terminal domain. D.A. Dougan, et al. **2002**, *FEBS Lett.* 529, 6–10; Fig. 1.

their partner proteins. Degradation of substrate membrane proteins occurs in a processive way, in which the protein is dislocated to the cytoplasm to become exposed to the proteolytic activity of FtsH. The primary criterion for membrane protein recognition is that they possess a cytoplasmic tail of sufficient length (≥ 20 amino acids).

Adaptor Proteins Expand or Modify the Substrate Specificity

A key question regarding ATP-dependent proteases concerns the nature and mechanism of their substrate specificity. ClpAP has been implicated in the degradation of a variety of different proteins, including the P1 phage encoded RepA and proteins with the proteolytic tag encoded by the *ssrA* RNA. In contrast, ClpXP has been shown to be involved in the degradation of λO protein, UmuD′ when present in an UmuD/D' heterodimer and SsrA-tagged proteins. But *in vivo*, only the ClpAP machine contributes to the degradation of proteins with the proteolytic tag.

In several cases, AAA^+ proteins utilize so-called *adaptor proteins* to modify or expand their substrate specificity. Adaptor proteins enhance the degradation of specific substrates without themselves being degraded. Six adaptor proteins have been described so far, five in *E. coli* and one in *B. subtilis*. ClpX cooperates with two such adaptor proteins, RssB and SspB. RssB is a factor specific for degradation of the starvation sigma factor σ^S, while SspB enhances the degradation of SsrA-tagged substrates. The third adaptor protein, ClpS, interacts with ClpAP and inhibits degradation of SssA-tagged proteins and ClpA itself. The UmuD′ protein is a component of DNA polymerase V, an error-prone polymerase that carries out translesion synthesis on damaged DNA templates (see Section 4.1.4). The intracellular concentration of the UmuD′ protein is strictly controlled by three different mechanisms: (a) regulated transcription, (b) post-translational processing of UmuD to UmuD′ and (c) degradation through ClpXP. UmuD functions as an adaptor molecule to deliver UmuD′ to the ClpXP protease. UmuD resembles SspB in that both proteins use related peptide motifs to bind to the N-terminal domain of ClpX. FtsH degrades cytoplasmic proteins as well as membrane-associated proteins. Two membrane proteins, HflK and HflC, redirect FtsH activity from membrane proteins such as SecY to cytoplasmic substrates such as σ^{32}. Both adaptor proteins require the periplasmic domain of FtsH. But how the activity of HflK and HflC is modulated *in vivo* to allow regulation of both membrane-associated and cytoplasmic proteins remains elusive. In *B. subtilis*, the adaptor protein MecA (or another related protein, YpbH) collaborates with ClpCP for the efficient degradation of specific proteins that regulate the development of competence. In summary, the joint use of tethering peptides and low-affinity primary degradation signal permits combinatorial control in regulated protein turnover.

While ClpS and ClpA form a complex, the other known adaptor proteins SspB, RssB and MecA interact directly with their substrate, thereby redirecting the substrate component to the AAA^+ component of the system. Furthermore, all adaptor proteins studied to date have been shown to positively regulate AAA^+ substrate binding. In contrast, ClpS, through interaction with ClpA, has distinct effects on

substrate binding, redirecting ClpA activity away from the degradation of SsrA-tagged proteins and towards degradation of a subset of specific aggregated or oligomeric proteins. Two possible molecular mechanisms of action by ClpS have been suggested: (a) ClpA undergoes a conformational change upon ClpS binding, blocking one class of binding site and revealing other(s) or (b) ClpA contains two separate substrate-binding regions, one of which is sterically blocked through ClpS binding. There are ~400 molecules of ClpS per cell and approximately 600 molecules of ClpA. One ClpA monomer binds one molecule of ClpS.

7.5.2 ATP-independent Proteases

The HtrA Family

HtrA (for *h*igh *t*emperature *r*equirement *A*), also called DegP, is a heat-shock protease localized in the periplasmic space of Gram-negative bacteria or integrated into the cytoplasmic membrane of Gram-positives. *E. coli* contains two additional periplasmic proteases belonging to the HtrA family, DegQ and DegS, where DegQ is a soluble and DegS a membrane-anchored proteases (Fig. 7.18). All three proteases are composed of an N-terminal domain believed to have regulatory functions, a trypsin-like protease domain (with the catalytic triad His105-Asp135-Ser210) and one or two PDZ domains (named after the three proteins *P*SD-95, *D*iscs-large, *Z*O-1). PZD domains mediate specific protein–protein interactions. Recent crystal structure analyses revealed that the *E. coli* HtrA forms a hexameric complex composed of two trimers. The basic trimer has a funnel-like shape with the protease domains located at its top and the PDZ domains protruding to the outside. A staggered association of two trimeric rings forms the functional HtrA hexamer, where the top and bottom of the molecular cage are constructed by the six protease domains, while the 12 PDZ domains generate the mobile sidewalls. The axial pores are completely blocked and the PDZ domains are the only gates allowing lateral access to the central cavity. Most importantly, the activity of HtrA can switch between chaperone and protease activities in a temperature-dependent manner. At temperatures below 28 °C, HtrA acts as a chaperone, protecting non-native proteins from irreversible aggregation. Above 28 °C, its protease activity increases dramatically, degrading non-native proteins. The activation of the proteo-

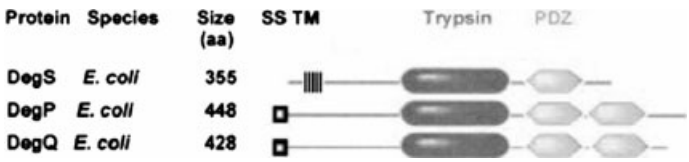


Fig. 7.18 Domain organization of the three Deg proteases of *E. coli*. While DegP and DegQ are soluble periplasmic enzymes which need a signal sequence (SS) to be translocated through the inner membrane, DegS is inserted into the cytoplasmic membrane via a transmembrane (TM) segment. The trypsin-like protease domain is followed by one or two PDZ domains. T. Clausen, et al. **2002**, *Mol. Cell* 10, 443–455; Fig. 1A.

lytic function of HtrA at elevated temperatures might occur by a conformational change, resulting in the opening of a helical lid to expose the active site.

HtrA plays an important role in the degradation of misfolded periplasmic proteins accumulated by heat shock or other stresses or after overproduction in the periplasm. In addition, it degrades hybrid and recombinant proteins, and mislocated proteins such as TreF and OmpF. HtrA is also involved in the pathogenesis of bacteria by degrading damaged proteins that are produced by reactive oxygen species released from the host defense system.

The RIP Family

RIP (*r*egulated *i*ntramembrane *p*roteolysis) is an important and widely conserved mechanism that governs signaling pathways in both pro- and eukaryotes. This mechanism involves the cleavage of a protein within a membrane or near the membrane surface. Cleaved products are released from the membrane and become biologically active by moving into the cytoplasm. The proteases that carry out RIP contain amino acid residues essential for catalysis, which are located in transmembrane segments and are designated intramembrane-cleaving proteases (I-Clips). RIP proteases share several structural features:

- They display similar hydrophobicity profiles predicting multiple membrane-spanning regions.
- They contain a putative zinc-binding motif, HExxH, at the amino-terminal end which is embedded in a hydrophobic environment.
- They contain a conserved motif, LDG (or NPDG) near the C-terminus, which is unique to the RIP family (function unknown).

Some RIB proteases have a PDZ domain, conserved structural elements of about 90 amino acids, which can occur in single or multiple repeated copies. They are involved in protein–protein interactions, and are used, e.g., in multimerization or protein targeting by signaling molecules.

The substrates of four different RIP endoproteases have been identified and analyzed. RseP of *E. coli* performs RIB of an anti-sigma factor, RseA, leading to activation of the σ^E regulon in response to unfolded proteins in the periplasm. RasP of *B. subtilis* performs RIB of the anti-sigma factor RsiW, activating the σ^W regulon in response to alkali shock. Pro-σ^K of *B. subtilis* is activated by RIP involving another endoprotease, SpoIVFB, in response to a signal from the forespore. Here, RIP removes its prosequence (amino acids 1–20), releasing σ^K from the outer forespore membrane into the mother cell cytoplasm. The last example is Eep from *E. faecalis*, which has been proposed to generate a peptide pheromone involved in mating *E. faecalis*.

The OmpT Protease

OmpT is a 33.5-kDa protease anchored in the outer membrane of *E. coli*. The enzyme cleaves peptides and proteins preferentially between two consecutive basic

amino acids. OmpT has been suggested to be involved in urinary tract disease, in DNA excision repair and in the breakdown of antimicrobial peptides, but the exact biological function remains to be elucidated. OmpT was classified as a member of a novel serine peptidase family called *omptin*. The Ser99 and His212 residues are essential active site residues.

The Tail-specific Protease

The *t*ail-specific *p*rotease (Tsp, also called Prc) represents a new class of proteases. It was first recognized as the periplasmic protease that cleaves the C-terminus of penicillin-binding protein 3 (PBP3). Later, it was identified as an activity that degrades a phage λ repressor variant with a nonpolar C-terminal sequence, WVAAA, but not the wild-type repressor, which has a polar C-terminal sequence, RSEYE. Tsp selectively degrades proteins with apolar amino acid residues in the last five positions at the C-terminus. This includes aberrant polypeptides synthesized from defective mRNAs using a peptide-tagging system. Cloning of the *tsp* gene revealed that it has no sequence homology to any of the well characterized members of other families and is not inhibited by several common protease inhibitors. Serine 430 and lysine 455, both located in the C-terminal portion of the protein, are essential for its catalytic activity. On the basis of this observation, it has been suggested that Tsp belongs to the family of proteases that use a serine–lysine dyad for catalysis. Members of this family include LexA and type-I signal peptidase. Tsp contains a PDZ domain as the substrate recognition domain, recognizing proteins with the proper C-terminal sequences for degradation.

Suggested Reading

M. Ehrmann, T. Clausen **2004**, Proteolysis as a regulatory mechanism, *Annu. Rev. Genet.* 38, 709–724.

S. Gottesman **2003**, Proteolysis in bacterial regulatory circuits, *Annu. Rev. Cell Dev. Biol.* 19, 565–587.

R.T. Sauer, et al. **2004**, Sculpting the proteome with AAA^+ proteases and disassembly machines, Cell 119, 9–18.

8
Secretion of Proteins

8.1
Introduction

All newly synthesized proteins need to be targeted to their site of action. While most proteins exert their function within the cytoplasm, extracytoplasmic proteins have to be transported into or across the cytoplasmic (inner) membrane, a process called *translocation*. In the Gram-negative bacteria, proteins can be inserted into the outer membrane or even secreted into the medium. Approximately 20% of the polypeptides synthesized by bacteria are located partially or completely outside of the cytoplasm. How does a protein know where to go? The information determining the subcellular localization site of a protein is encoded in its amino acid sequence in most cases. The localization information is usually presented as a short sequence segment called a *protein sorting signal*.

Due to the cell envelope structure, Gram-positive bacteria have three distinct protein localization sites where proteins are secreted: the cytoplasmic membrane, the cell wall and the exterior space of the cell. Similarly, Gram-negative bacteria even have five localization sites: the cytoplasm, the inner membrane, the periplasm, the outer membrane and the exterior space. In Gram-positive bacteria, a protein will be secreted to the outside if it has a signal peptide at its N-terminus and it will be integrated into the cytoplasmic membrane if is has a transmembrane segment(s). In addition, there is a *sorting pathway* for the cell wall proteins. In Gram-negative bacteria, a protein will pass through the inner membrane if it has a signal peptide but does not have an additional transmembrane segment(s); if it has transmembrane segments, it will be integrated into the inner membrane. The sorting mechanism between the periplasm, the outer membrane and the outside medium is not yet fully understood, but a signal for the outer membrane proteins has been proposed. Both types of bacteria have *lipoproteins*, i.e., proteins that have a covalently linked lipid moiety. Lipoproteins have a slightly different type of signal peptide and in Gram-negative bacteria they are further sorted to either the inner or the outer membrane. In addition, Gram-negative bacteria have developed six different specific secretion systems. The expression *secretom* has been coined to denominate all the proteins secreted by a given species into the extracellular milieu.

Dynamics of the Bacterial Chromosome. Structure and Function. W. Schumann
Copyright © 2006 Wiley-VCH Verlag GmbH & Co. KGaA
ISBN 3-527-30496-7

8.2
Protein Sorting Signals for Subcellular Localization

Four different protein sorting signals have been described: *signal sequences*, *transmembrane sequences*, *lipoprotein sorting signal* and *cell wall sorting sequences*. While signal sequences are important for translocation of proteins through the cytoplasmic membrane and are cleaved off during translocation by a *signal peptidase*, transmembrane sequences allow the insertion of proteins into membranes and cell wall sorting sequences are required to anchor proteins into the cell wall of Gram-positive cells.

Signal Sequences

Signal sequences (also called leader sequences or signal peptides) are the amino-terminal extension of polypeptides that direct them to and across the cytoplasmic membrane in prokaryotes (and the endoplasmic reticulum in eukaryotes). Alignment of hundreds of different signal sequences has revealed that a concrete consensus sequence does not occur in signal peptides; rather a three-region structure is conserved, the N-, H- and C-regions. The most amino-terminal N-region often contains positively charged residues, i.e., arginine(s) or lysine(s), and has been suggested to interact with the translocation machinery and negatively charged phospholipids in the lipid bilayer of the membrane during translocation. The central H-region ranges from seven to 15 residues and is formed by a stretch of ten and more hydrophobic residues that seem to adopt an α-helical conformation in the membrane. Helix-breaking glycine or proline residues are frequently found in the middle of the hydrophobic core. The carboxyl-terminal C-region contains more polar residues than the H-region and there is a weak consensus pattern specifying the cleavage site for the signal peptidase. The mature part of the translocated protein is released from the membrane and can fold into its native conformation. Finally, the signal peptide is degraded by *signal peptide peptidases*.

At present, four major classes of amino-terminal signal peptides can be distinguished on the basis of the signal sequence recognition sequence, as shown in Fig. 8.1: Sec-type, twin-arginine type, lipoprotein, prepilin-like, bacteriocin and pheromone signal peptides. The Sec-type and the twin-arginine signal peptide (Fig. 8.1a) is recognized by the *signal peptidase I* (one enzyme in *E. coli*, five different enzymes in *B. subtilis*), where the latter signal peptide is distinguished from the Sec-type by the conserved SRRxFLK motif. The C-region of both signal peptides ends with the consensus sequence A-X-A at position –3 to –1, relative to the signal peptide. Lipoprotein signal peptides are shorter than those of the Sec- and twin-arginine types. The C-region contains the *lipobox* (consensus sequence L-(A/S)-(A/G)-C, of which the invariable cysteine residue is the target for lipid modification and becomes the first residue of the mature lipoprotein) and is cleaved off by the *signal peptidase II* (Fig. 8.1b). Prepilin-like signal sequences contain the C-region with the cleavage site localized between the N- and H-regions; and the *prepilin signal peptidase* acts at the cytoplasmic side of the membrane (Fig. 8.1c).

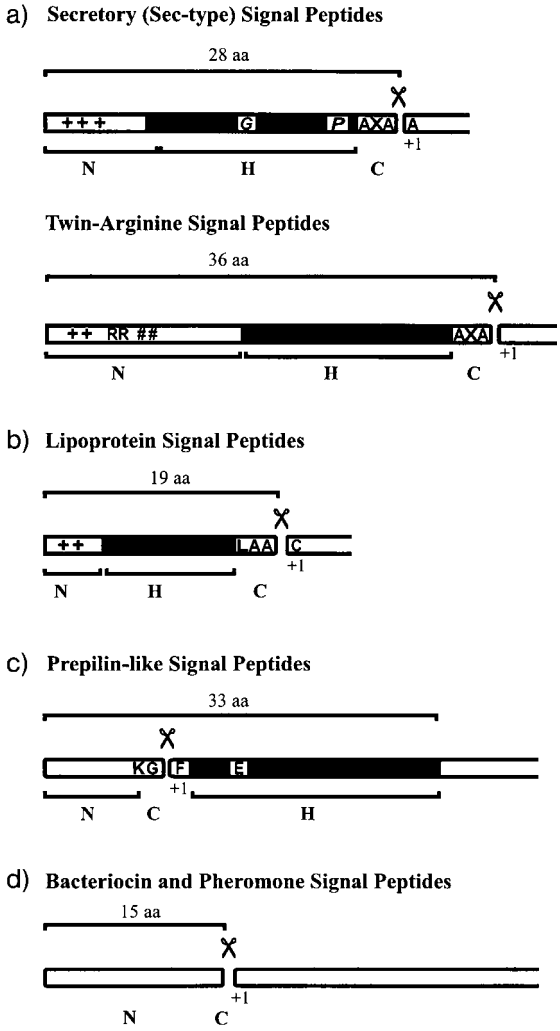


Fig. 8.1 Four different signal sequences allow proteins to translocate through the inner membrane. (A) Secretory and twin-arginine signal peptides consist of three regions designated as N-, H- and C-region where the N-region contains positively charged, the H-region hydrophobic and the C-region more polar amino acid residues and a weak consensus pattern (AXA↓A) specifying the cleavage site for signal peptidase I. The twin-arginine signal peptide contains two consecutive arginine residues followed by two hydrophobic residues as part of the N-region. (B) Signal sequences of lipoproteins are shorter and carry a lipobox sequence (LAA↓C) at their end recognized by signal peptidase II. (C) Prepilin-like signal peptides are even shorter, and cleavage by the prepilin signal peptidase occurs within the cytoplasm. (D) Signal peptides of bacteriocins and pheromones consist only of N- and C-regions and are translocated via ABC transporters. The mean length for all four different signal peptides for *B. subtilis* is indicated above the signal sequence. H. Tjalsma, et al. **2004**, *Microbiol. Mol. Biol. Rev.* 68, 207–233; Fig. 2.

Here, the hydrophobic H-region remains attached to the mature protein. The last class of signal peptides present on bacteriocins and pheromones translocated via ABC transporters (see Section 1.2) consists of only N- and C-regions and completely lack the hydrophobic H-region (Fig. 8.1d).

Transmembrane Sequences

Integral membrane proteins are found in the cytoplasmic membrane and additionally in the outer membrane of Gram-negative bacteria. They are integrated into the membrane via one or more hydrophobic regions called *transmembrane sequences*. They carry a signal sequence-like element called the *signal anchor sequence* at their N-terminus that is topologically similar to a noncleavable signal sequence and assumes an N-in, C-out transmembrane orientation. One or more hydrophobic regions (*stop-transfer sequences*) that follow a signal or signal anchor sequence halt translocation, thereby anchoring the protein within the inner membrane. These proteins exhibit a wide variety of different topological states and are divided into three classes: *monotopic, bitopic* and *polytopic* proteins (Fig. 8.2). Monotopic proteins are only partially embedded in the lipid bilayer, while bitopic proteins have a single transmembrane segment and polytopic proteins have at least two. Bitopic membrane proteins can be subdivided into three subclasses depending on their orientation. Type Ia and Ib have their C-terminus exposed to the cytoplasm, whereas that of type II membrane proteins is on the outside. Type Ia bitopic proteins are synthesized with a signal peptide which is cleaved off by the signal peptidase. A second hydrophobic region anchors them in the cytoplasmic membrane. Integration of all the other inner membrane proteins occurs independently of any signal sequence. Type Ib membrane proteins have a single stretch of hydrophobic amino acids at their extreme N-terminus; and this transmembrane sequence anchors them in the lipid bilayer. Type II proteins differ from type Ib proteins only in the orientation in which they insert; and the transmembrane sequence is usually located close to the N-terminus, as shown in Fig. 8.2. Polytopic inner mem-

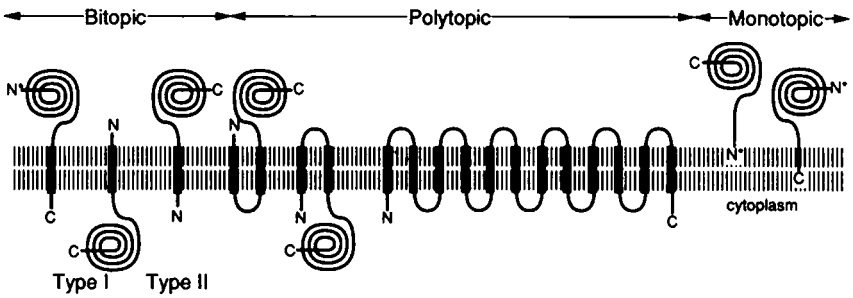


Fig. 8.2 Topologies of integral membrane proteins. Monotopic proteins are inserted into the inner membrane with either their N- or C-terminus. Bitopic proteins cross the membrane completely in such a way that both ends are separated by the lipid bilayer. Polytopic proteins, in contrast, cross the membrane at least twice. A.P. Pugsley **1993**, *Microbiol. Rev.* 57, 50–108; Fig. 12.

brane proteins span the membrane at least twice via segments of mainly hydrophobic amino acids.

The transmembrane sequence of the majority of membrane proteins with known three-dimensional structures are α-helices comprised of apolar residues. In another class of membrane proteins, all the transmembrane sequences are comprised of β-helices; and this type of protein has been discovered in the outer membrane proteins of Gram-negative bacteria.

Lipoprotein Sorting Sequence

Lipoproteins present in various bacteria are synthesized as precursors in the cytoplasm and then translocated across the inner membrane. Subsequent modification reactions sequentially take place on the periplasmic side of the inner membrane, leading to the formation of mature lipoproteins having a lipid-modified cysteine at the N terminus. The murein liporotein of *E. coli* was the first example of a lipid-modified protein in which the structure of a covalently bound lipid and its linkage to a protein was elucidated. This protein exists in two forms, one of which is bound to the peptidoglycan and the other is free in the outer membrane (see Section 1.5). Lipoprotein is synthesized as a preprotein with an about 20-amino-acid signal sequence at the N-terminus. The first amino acid after the signal peptide is a cysteine residue; and this will become the N-terminal residue of the mature protein. Modification of the cysteine residue occurs in four steps, involving three different enzymes and resulting in *N*-acyldiacylglyceryl-cysteine (Fig. 8.3):

1. While the lipoprotein is still attached to its signal peptide, the first step is the transfer of the nonacylated glyceryl moiety of phosphatidylglycerol to the sulfhydryl group of the prospective N-terminal cysteine residue by prolipoprotein–phosphatidylglycerol glyceryl transferase, encoded by the *lgt* gene.
2. Next, the *sn*-3 and *sn*-2 hydroxyl groups of the glyceryl moiety are acylated by long-chain fatty acids exchange from phospholipids by the enzyme phospholipid transacylase (gene *lnt*). This diacylglyceryl modification is a prerequisite for the cleavage of the signal peptide by the lipoprotein-specific signal peptidase.
3. Then, the prelipoprotein is translocated through the inner membrane with removal of the signal peptide by signal peptidase II (gene *lsp*).

In the last step, the N-terminal cysteine moiety is further *N*-acylated by *N*-acyltransferase, using phosphatidyl glycerol.

Binding of lipoprotein to the peptidoglycan occurs through a peptide bond between the C-terminal lysine and the carboxyl group of the peptide moiety of the peptidoglycan.

In *E. coli*, lipoproteins are anchored not only to the outer but also to the inner membrane in both cases on the periplasmic side through the N-terminal lipids. The membrane specificity of lipoprotein localization is determined by the residue next to the modified cysteine residue. Lipoproteins possessing aspartic acid at this

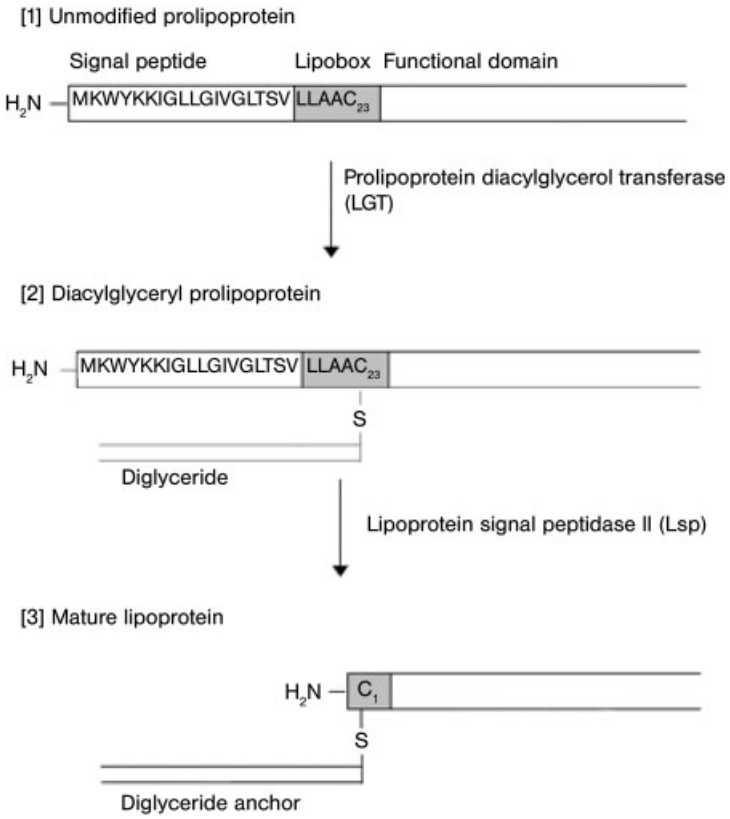


Fig. 8.3 Lipoprotein biosynthesis in Gram-positive bacteria. First, the lipobox cysteine is modified by a diglyceride through a thioester linkage. Second, signal peptidase II removes the signal peptide. In low G+C bacteria, an additional fatty acid is amide-linked to the free amino-terminus (not shown). I.C. Sutcliffe, D.J. Harringdon **2002**, *Microbiology* 148, 2065; Fig. 1.

position are localized at the inner membrane, whereas those possessing amino acids other than aspartic acid are destined for the outer membrane (see below).

Cell Wall Sorting Sequences

Gram-positive bacteria mostly code for two enzymes called *sortases* which are involved in the covalent anchoring of proteins in the cell wall. These proteins are recognized by the sortases by virtue of a *sorting sequence* present at the C-terminus of these proteins. This sorting sequence is composed of three parts: a *sorting motif* followed by an amino acid sequence enriched in hydrophobic amino acid residues and a few positively charged amino acids at the immediate C-terminus. The total length of the sorting sequence comprises 30–35 amino acid residues. Two sorting motifs have been identified so far in the proteins of *S. aureus* and other Gram-po-

sitive species. While the motif LPXTG is recognized by sortase A, the motif NPQTN is recognized by sortase B. The mechanism for anchoring proteins on the cell wall is described in detail under Section 8.6.

8.3 Cytosolic Targeting Factors

The Sec translocation machinery, unlike the Tat machinery, is unable to handle tightly folded proteins. A protein may be maintained in an *translocation-competent* state in several different ways.

1. The protein may be translocated across the inner membrane simultaneously with translation of the transcript (cotranslational translocation), thus ensuring that not even its secondary structures are formed in the cytoplasm.
2. Molecular chaperones or antifolding factors may prevent folding in the cytoplasm.
3. In some cases, signal sequences act as intrapolypeptide chaperones to prevent rapid folding.
4. Proteins that are exported post-translationally may have disulfide bonds in their final structures that do not form in the cytoplasm so that the proteins cannot attain their final conformations in the cytoplasm. Therefore, post-translational translocation of a protein involves a cytosolic intermediate of the precursor which has to be stabilized in a more or less unfolded conformation designated as the *translocation-competent form* and targeted to the inner membrane. This stabilization and targeting of proteins to the membrane is mediated by two pathways exhibiting different substrate specificity. While large hydrophobic membrane proteins are selectively recognized by the bacterial *signal recognition particle* (SRP), the translocation of secretory proteins involves the post-translational binding of most of the preproteins by the *molecular chaperone* SecB in *E. coli*. The SRP has a preference for signal sequences with higher hydrophobicity.

Molecular Chaperones

Many Gram-negative organisms, including *E. coli*, contain a secretion-dedicated chaperone called SecB, already described under Section 7.2.2. This chaperone functions by rapidly binding to partially folded precursors either free in the cytosol or bound to the ribosome as nascent chains. Next, the binary SecB-precursor protein is targeted to SecA. Besides SecB, some SecB-independent proteins, such as β-lactamase, alkaline phosphatase and ribosome-binding protein, are kept in a translocation-competent form by the general chaperone machines DnaK or GroEL (see Section 7.2.1).

Signal Recognition Particle

The translocation of precursor proteins across the endoplasmic reticulum of eukaryotes depends on the signal recognition particle (SRP). The mammalian SRP complex consists of six proteins that associate with an RNA molecule (7S RNA). As a nascent polypeptide chain emerges from the ribosome, its amino-terminal sequence is recognized by the 54-kDa protein of the SRP (SRP54) and the binding of SRP causes a translation elongation stop. The SRP ribosome nascent chain complex is then delivered to the SRP receptor at the endoplasmic reticulum membrane, where the nascent chain is released to the membrane-embedded translocon and subsequently translocated into the endoplasmic reticulum.

In eubacteria, homologs of some of the SRP components have been identified, representing a simplified version of the mammalian counterpart. Instead of six proteins and one RNA, the *E. coli* SRP consists only of a 48-kDa protein, Ffh (for *f*ifty-*f*our *h*omolog), and the 4.5S RNA (encoded by the *ffs* gene). The mature 4.5S RNA is 114 nucleotides long, much smaller than its mammalian homolog 7S RNA (over 300 nucleotides long). The Ffh protein exhibits a well defined domain structure and consists of a G-domain at its N-terminal part, which contains three GTP-binding motifs and has been shown to interact with FtsY, the SRP receptor homolog. The C-terminal M-domain of Ffh is rich in methinine residues and contains three amphipathic helices. The first of these helices is situated close to the G-domain and is responsible for the specific binding of Ffh to precursors of secreted proteins. The other two helices are in close vicinity to the C-terminus and bind the RNA. The *E. coli* SRP receptor, FtsY, unlike the heterodimeric eukaryotic SRP receptor, consists of only one subunit. The *ffh*, *ffs* and *ftsY* genes, which encode Ffh, 4.5S RNA and FtsY, respectively, are all essential for cell viability.

8.4
Translocases Move Newly Made Proteins Into or Across Membranes

Each membrane is made up of a specific and unique set of proteins, which determines its identity. As a consequence, the insertion of proteins cannot be a spontaneous event. Indeed, insertion of proteins into or across the membrane depends on a set of proteins called *translocases*, which is sited in the membrane. In Gram-negative bacteria, seven translocases have been identified so far, three in the inner and four in the outer membrane. The three translocases present in the inner membrane are the *SecYEG pathway*, the *Tat pathway* and *YidC*. YidC works either independently or together with SecYEG to insert proteins into the inner membrane. Most proteins to be translocated through the cytoplasmic membrane use the Sec pathway; and these proteins are almost unstructured. In contrast, proteins to be translocated by the Tat pathway must completely fold in the cytoplasm, where most of them bind cofactors.

8.4.1 The Sec Pathway of Protein Translocation Through the Inner Membrane

Most bacterial proteins are translocated across the inner membrane through the Sec pathway, therefore also called the *general secretion pathway*. These include the periplasmic and outer membrane proteins in Gram-negative bacteria which are synthesized with an N-terminal signal sequence and which are targeted to the inner membrane by the molecular chaperone SecB. This pathway consists of four components, the targeting factor, the SecA motor protein, the SecYEG translocon forming a pore within the inner membrane and the accessory heterotrimeric SecDFYajC complex.

The SecA Motor Protein

SecA is a precursor protein-stimulated ATPase that performs a central role in bacterial protein translocation, as it acts as the molecular motor protein. SecA is a homodimer with a subunit mass of about 100 kDa. It functions as a motor protein (translocase) that couples the energy of ATP binding and hydrolysis to the stepwise translocation of precursor proteins. SecA binds precursor proteins through their signal peptide and mature domain; and it exhibits a low endogenous ATPase activity, which can be stimulated by adding membranes, the SecYEG complex and precursor proteins. It penetrates through the center of the SecY channel or, alternatively, into the lipid bilayer, to contact the translocating polypeptide chain. The ATPase drives presecretory proteins through the translocation channel in batches of approximately 25 residues at a time. The SecA protomer is composed of an N-terminal ATPase domain and a C-terminal domain with a lipid-binding site as well as a Zn^{2+}-coordinating SecB-binding site. The ATPase domain contains two *n*ucleotide *b*inding *d*omain folds, NBD1 and NBD2, for high- and low-affinity binding, respectively, as well as a preprotein-binding site. Whereas SecA has a low intrinsic ATPase activity, it is strikingly activated under the translocation conditions in the presence of preprotein and SecYEG. While NBD1 ($K_{d(ADP)} \approx 0.15$ μM) is solely responsible for the catalysis, NBD2 ($K_{d(ADP)} \approx 340$ μM) functions as an intramolecular regulator of ATP hydrolysis at NBD1. The intrinsic ATPase activity is downregulated by two mechanisms where the C-terminal domain and the NBD2 region inhibit the activity independently. Interaction between a preprotein and the membrane liberates SecA from this downregulation. The rate-limiting step in overall ATP hydrolysis is the release of ADP for its exchange with ATP.

SecA appears to be the only *sec* gene that is under protein secretion-specific regulation where *secA* is up-regulated under conditions of lowered functionality of the Sec translocation machinery. Inhibition of protein secretion by either genetic or biochemical means leads to 10-fold induction of *secA* translation. Analysis of this system revealed that *secA* is the second gene in a *secM-secA-mutT* operon (Fig. 8.4) and that translation of *secA* is coupled to the translation of *secM* (for *sec*retion *m*onitor). The SecM protein contains an arrest sequence ($F^{150}X_4WIX_4$

GIRAGP[166]), which interacts with the ribosomal exit tunnel to halt translation elongation beyond Pro-166. This elongation arrest is transient under normal conditions, where the nascent SecM precursor interacts with the SRP-Sec translocation system. In contrast, the arrest is prolonged strikingly in the absence of active export of SecM, due either to a *cis*-defect in its signal sequence or a *trans*-defect in the Sec machinery. The stalled ribosome may then disrupt the secondary structure of the *secM* mRNA, leading to the exposure of the *secA* translation initiation sequence for the entry of new ribosomes that translate *secA* (Fig. 8.4). In summary, the arrest sequence of SecM has at least two functions in SecA translation:

1. The transient elongation arrest in normal cells is required for the synthesis of SecA at levels sufficient to support cell growth.
2. The prolonged SecM elongation arrest under conditions of unfavorable protein secretion is required for the enhanced expresion of SecA to cope with such conditions.

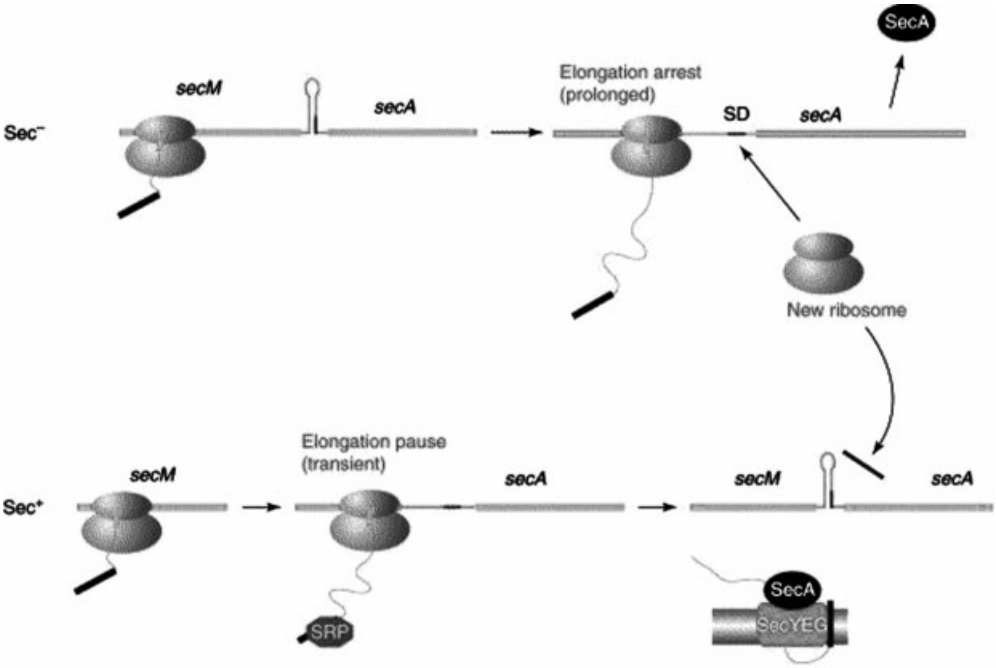


Fig. 8.4 Translation of the *secA* transcript is regulated by elongation arrest involving the upstream *secM* reading frame. Under conditions of unsufficient SecA within the cytoplasm (depicted as Sec^-), ribosomes translating the *secM* gene are stalled (SecM itself is translocated to the periplasm through the Sec system) preventing the formation of a stem-loop structure (which would sequester the Shine–Dalgarno sequence for *secA*) between *secM* and *secA*. This in turn allows binding of ribosomes to the Shine–Dalgarno sequence preceding *secA* and successful translation of this gene. If sufficient amounts of SecA are present (Sec^+), ribosomes translate *secM* and dissociate, allowing formation of the stem-loop structure reducing translation of *secA*. H. Mori **2002**, *Trends Microbiol.* 9, 494–500; Fig. 3.

The SecYEG Translocon

The proteins coding for the three genes *secY*, *secE* and *secG* form an aqueous channel called a *translocon*. The polytopic SecY protein is the largest subunit and contains ten transmembrane segments. SecE is a small integral membrane protein, forms three transmembrane segments and associates with SecY, thereby preventing SecY from being degraded by FtsH, an ATP-dependent protease anchored in the cytoplasmic membrane (see Section 7.5.1). In *E. coli*, uncomplexed SecY is rapidly degraded by FtsH. SecG, the third component of the translocase, is not required for translocation of proteins, but it contributes to the efficiency of the reaction. It shows the unusual property of inverting its orientation in the membrane. This occurs under conditions that would stabilize the inserted status of SecA. In *E. coli*, pore formation requires the activation of SecA by precursor protein and ATP. Under these conditions, the SecA dimer recruits four SecYEG heterotrimers to form a superassembly with a large central opening of 20–30 Å that may function as the protein-conducting channel.

Three-dimensional X-ray crystallographic analysis (at 3.2 Å resolution) of the Sec complex of the archaeon *M. jannashii* represents a major step forward in understanding protein translocation. The crystal structure of SecY reveals ten transmembrane segments; and the steeply angled SecE transmembrane helix forms a girdle around part of the SecY barrel, thereby stabilizing SecY. The crossed organization of SecY transmembrane segments 2 and 7 could form a lateral opening, through which signal peptides (inserted laterally into the translocase and transmembrane segments of integral membrane proteins) escape from the lumen of the channel into the lipids of the membrane.

The SecD-SecF-YajC Complex

SecD, SecF and YajC, three integral membrane proteins, form a heterotrimeric complex in *E. coli* which loosely associates with SecYEG to form a supramolecular translocase complex. In *B. subtilis* and some other organisms, SecD and SecF are fused into a single membrane protein. The role of this complex remains elusive. The following functions have been suggested: (a) modulation of the catalytic cycle of the SecA protein, (b) maintenance of the proton motive force (PMF), (c) assembly of the translocase, (d) prevention of backward sliding of translocated proteins and (e) clearing the translocation channel of signal peptides or misfolded proteins. Simultaneous depletion of all three proteins results in a very severe inhibition of protein translocation.

The Sec or General Translocation Pathway

Proteins transported via the Sec pathway cross the cytoplasmic membrane through the SecYEG channel; and two alternative modes of translocation have been identified providing the driving force. In *co-translational translocation*, as the signal sequence emerges from a translating ribosome, it binds to the SRP and, in concert with its membrane-bound FtsY, the nascent polypeptide chain is targeted

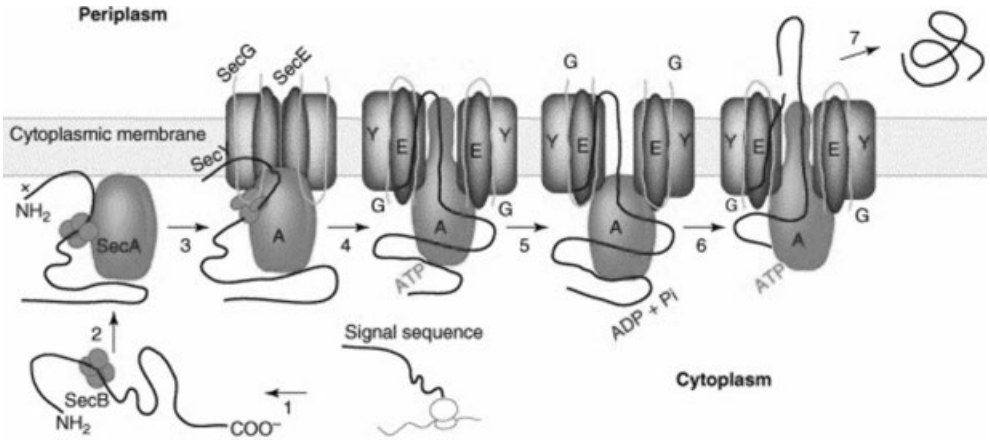


Fig. 8.5 The Sec or general translocation process. Most nascent polypeptide chains to be translocated are recognized and bound by the SecB chaperone (step 1) which targets it to a SecA homodimer (step 2). Upon dissociation of SecB, SecA will bind to the inner membrane causing SecYEG to form the translocon (step 3). Next, SecA with the bound preprotein inserts into the channel (step 4) which requires bound ATP. Upon hydrolysis of ATP, SecA retracts (step 5), binds to another segment of the protein which is feeded together with SecA into the channel (step 6). During each translocation step, 20–30 amino acid residues are pushed through the translocon. H. Mori **2002**, *Trends Microbiol.* 9, 494–500; Fig. 1. (This figure also appears with the color plates.)

to the SecYEG channel. The energy for translocation comes from GTP hydrolysis during translation. In the *post-translational translocation*, many newly synthesized precursors interact with the chaperone SecB (see Section 7.2.2) that targets it to the Sec translocase by its ability to bind to SecA (Fig. 8.5). The preprotein-bearing SecA then binds to the membrane, at a high-affinity SecA-binding site made up of SecY and SecE, followed by initiation, elongation and termination of translocation, either all the way to the extracytosolic side or into the lipid phase of the membrane. During initiation, the signal peptide sequence inserts into the membrane in a loop-like configuration, leaving the N-terminus on the cytosolic side. At the end of the initiation phase, the signal peptide cleavage site has reached the periplasmic surface and approximately 20–30 mature residues are already within the membrane. Elongation of translocation takes place in a stepwise fashion with a step of 20–30 amino acid residues, involves SecA insertion–deinsertion cycles and requires cycles of ATP hydrolysis and/or proton motive force across the membrane. Little is known about the termination process, which occurs on the periplasmic side, leading to the release and/or folding of the substrate protein into the periplasmic space. It has been suggested that SecDF and SecY are involved in the late-step reactions.

8.4.2
The Tat Pathway of Secretion Proteins Through the Inner Membrane

In bacteria, the vast majority of extracytoplasmic proteins are transported across the cytoplasmic membrane via the Sec pathway while they are unfolded. In contrast, the physiological role of the Tat pathway is to transport folded proteins. The prime role of the Tat system appears in the transport of proteins, which are either obliged to fold prior to translocation, or which fold too tightly for the Sec system to accommodate. Examples of the former category include periplasmic proteins that are exported only after binding a range of complex redox cofactors, such as FeS centers or molybdopterin. These cofactors are inserted in the cytoplasm by complex enzymatic processes. The Tat pathway is distinct from the Sec one in that: (a) Tat substrates are secreted in a folded conformation, (b) Tat signal peptides contain a highly conserved twin-arginine motif, (c) the energy-driving translocation is provided exclusively by the proton motif force and (d) the Tat pathway is not universally conserved.

Tat substrate proteins are synthesized with N-terminal signal sequences containing a consensus SRRxFLK motif in which the arginine residues are invariant. These 'twin arginine' signal peptides target precursor protein to the recently discovered twin arginine translocation (Tat) protein export system which is mechanistically and structurally related to the ΔpH-dependent thylakoid import pathway of chloroplasts. Remarkably, both bacterial and plant Tat systems have been shown to transport fully folded proteins, using energy provided by the transmembrane electrochemical gradient.

The components of this translocation system consist of four membrane proteins: TatA, TatB, TatC (whose genes are arranged in an operon together with *tatD*, which is not required for protein translocation) and TatE, encoded by an unlinked gene. TatA, TatB and TatC are all required for protein translocation and are estimated to be in the cell at a molar ratio of 40:2:1. TatA and TatE are homologous proteins with overlapping functions in the Tat pathway. The *tatC* gene product is predicted to contain four or six transmembrane helices, while the other three contain only one. It is assumed that at least some of these proteins form a translocation channel which has to be firmly closed in the absence of proteins to be translocated, but the exact nature of this channel remains elusive. Mutations in any of the genes that encode these components give rise to the inability to process twin-arginine-containing proteins efficiently. The core Tat transporter is a large oligomeric complex of the TatA, TatB and TatC proteins. The TatBC unit is believed to form the twin-arginine signal peptide recognition module, while TatA forms a very large oligomeric ring structure, presumed to be the protein-conducting channel itself. This aqueous channel must open wide enough to accommodate substrates having a diameter of up to 70 Å. The number of α-helices required to enclose an aqueous channel of this diameter is greater than 20. Given the relatively small size of the Tat proteins, multiple copies of at least one of the Tat proteins would be required to form that channel. Since TatA is present in large molar excess over the other Tat proteins, it is an obvious candidate for the channel-forming component.

The Tat system is not ubiquitous. It is present in all bacterial genomes whose complete sequence has been determined bigger than 2 Mbp, but is absent from the two genomes smaller than 1 Mbp. It is also absent from the half of bacterial genomes between 1 Mbp and 2 Mbp. The minimum components consist of TatC and TatA proteins for bacteria. Most interestingly, almost all secreted proteins from halophilic Archaea are putative Tat substrates. It has been hypothesized that the extensive use of the Tat pathway by the Halobacteriaceae represents an evolutionary solution to the problem posed by the need for stabilization of secreted proteins in high salt conditions. Quite recently, it has been shown that the Tat pathway can also associate in *P. aeruginosa* with the type II secretion system, allowing secretion of proteins into the extracellular milieu (see Section 8.6.3).

What prevents premature targeting of cofactor-containing proteins before cofactor loading is complete; and what coordinates the export of of multiprotein complexes? Proofreading chaperones bind to the twin-arginine signal peptide, and in some cases, specific or general chaperones also bind to the mature part of the protein. In *E. coli*, two such specific chaperones are TorD (with TorA) and HybE (with hydrogenase-2).

A protein that specifically binds the twin-arginine signal sequence of the *E. coli* dimethylsulfoxide (Me2SO) reductase (DmsA subunit) has been discovered. This 204-residue protein, DmsD, has homology to the TorD family of molecular chaperones. Sequence analysis predicts that members of this family comprise at least two distinct structural domains. DmsD has also been shown to interact with the precursor form of trimethyl *N*-oxide (TMAO) reductase (TorA), a DmsA homolog that also binds a molybdopterin cofactor. However, DmsD is unable to interact with the fully folded mature forms of DmsA and TorA, suggesting that it interacts with the twin-arginine signal sequence. Both proteins are members of the TorD family of molecular chaperones. While these two proteins are apparently associated with the biosynthesis of DmsA-homologous molybdoenzymes, other families of Tat-targeted metalloproteins may have analogous cofactor chaperones.

Recently, it has been shown that, of the approximately 26 Tat substrates in *E. coli*, at least five are genuine inner membrane proteins. They possess single transmembrane α-helices at the extreme C-terminus, where the N terminus is exposed in the periplasm and the C terminus in the cytoplasm.

Key Tat-dependent components of the respiratory chain, (NiFe) hydrogenase and trimethylamine *N*-oxide (TMAO) reductase, have been shown to acquire their redox cofactors and even oligomerize prior to the transport event.

In *B. subtilis*, there are two different TatA (TatAd, TatAy) and TatC (TatCd, TatCy) proteins where TatAdTatCd and TatAyTatCy both function as individual, substrate-specific translocases for the twin-arginine precursors proteins PhoD and YwbN, respectively. These minimal TatAC translocases of *B. subtilis* are representative for the Tat machinery of the vast majority of Gram-positive bacteria, *Streptomycetes* being the only known exception with TatABC translocases.

T. Palmer, B.C. Berks **2003**, Moving proteins across the bacterial cell membrane, *Microbiology* 149, 547–556.

8.4.3 Insertion Proteins Into the Inner Membrane

Evidence is accumulating that, in *E. coli*, inner membrane proteins can be targeted and inserted into the inner membrane via different pathways. The best characterized inner membrane protein assembly pathway is the SRP/Sec-translocase pathway. But when *E. coli* cells are completely depleted of SRP, insertion of proteins is not completely blocked, suggesting that alternative pathways must occur. Indeed, three different integration routes seem to operate in *E. coli*, one involving the Sec translocon, a second YidC and a third SecYEG and YidC.

The Sec Pathway

The mechanism of transfer of the nascent chain from the SRP to the translocon is still enigmatic, but could involve direct interactions of FtsY and SRP with the translocon. Alternatively, it is possible that the membrane surrounding the translocon is enriched in acidic phospholipids that bind FtsY. As outlined above, SecA triggers oligomerization of the SecYEG heterotrimers. In the present case, this is caused either by SecA too, (insertion of FtsQ seems to dependend on SecA) or, if SecA is not involved, either by the SRP or by the ribosomes. The translocon scans polypetide chains for hydrophobic domains and stops translocation if a hydrophobic stretch is present. Next, such transmembrane segments have to move from the translocon into the lipid bilayer. This could occur either spontaneously or could be triggered by YidC, depending on the protein substrate.

The YidC Pathway

The 61-kDa YidC translocase is located in the cytoplasmic membrane in both Gram-negative and Gram-positive bacteria and is involved in the insertion of a subset of integral membrane proteins into the inner membrane, but not in the translocation of exported proteins. Possibly, the function of YidC is to recognize hydrophobic regions of a membrane protein and to catalyze the integration of these regions in a transmembrane orientation into the membrane bilayer. YidC is a polytopic integral membrane protein that spans the membrane six times with an N-in, C-in topology and a large, poorly conserved 319-residue periplasmic domain between transmembrane segment one (TM1; signal anchor sequence) and TM2. YidC is targeted cotranslationally by the SRP to the Sec-translocon, which appears to be required for proper assembly. Depending on the substrate protein, YidC works either together with the Sec translocase or on its own. Membrane insertion of some Sec-dependent proteins such as FtsQ, mannitol permease and the signal peptidase I is affected to some extent by the decrease in YidC expression. YidC has been shown to interact with these nascent polypeptides as they move laterally from the Sec-translocon into the lipid bilayer. This interaction appears to be specific for the transmembrane segments in the nascent polypeptide. Moreover, YidC could be copurified with the Sec-translocon, suggesting a physical connec-

tion. Some bacterial membrane proteins have been shown to insert independently of the Sec translocase. These are the single- or double-spanning coat proteins of the P3 and M13 filamentous bacteriophages, which have small periplasmic regions. Whereas the single-spanning PF3 coat protein comprises only 44 amino acids, the 73-amino-acid M13 procoat protein is synthesized with a leader peptide and is inserted as a double-spanning protein. Both these proteins require YidC for their insertion and the translocation of their periplasmic region across the membrane.

In summary, YidC seems to fulfill two functions during the translocation and insertion process of inner membrane proteins. With some proteins, it acts independently of any other component; and with other proteins, it acts in collaboration with the Sec translocon. It remains an open question as to which features of a protein determine its specificity. Furthermore, it is a tempting proposal that YidC is a factor involved in lateral movement from the translocase to the lipid phase.

8.5 Anchoring of Lipoproteins Into the Outer Membrane

E. coli possesses more than 90 lipoproteins, most of which are anchored to the outer membrane through N-terminal lipids. All these proteins are thought to be synthesized with a signal peptide at the N-terminus and then be translocated across the cytoplasmic membrane in a Sec machinery-dependent manner. Cysteine residues at the N-termini of mature lipoproteins are modified with lipids, which function as an anchor to the membrane. Lipid modification and processing to mature lipoproteins in *E. coli* have been shown to take place in the inner membrane, followed by their localization to either the inner or outer membrane. While anchoring into the inner membrane occurs spontaneously, outer membrane lipoproteins are escorted to their final destination involving five different proteins. As already mentioned, the amino acid residue next to the N-terminal cysteine acts as the discriminator but is also affected by the residue at position 3. Thus, strong inner membrane retension occurs with aspartate at position 2 and with aspartate, glutamic acid, glutamine or aspartate at position 3. If the residue at position 2 is not an aspartate, the ATP-binding cassette (ABC) transporter, LolCDE, located in and at the inner membrane (Fig. 8.6), releases outer membrane-directed lipoproteins from the inner membrane in an ATP-dependent manner. In this way, aspartate at position 2 (in combination with certain residues at position 3) functions as a LolCDE avoidance signal. Next, the lipoprotein forms a complex with LolA. The lipoprotein–LolA complex crosses the periplasm and then interacts with the outer membrane receptor LolB, which is essential for the anchoring of lipoproteins to the outer membrane (Fig. 8.6). Lipoproteins such as Lpp are transferred from LolA to LolB. The Lpp-LolB complex most probably represents an intermediate of the outer membrane localization. While the number of Lpp molecules in E. coli cells exceeds 10^5, that of LolB and LolA is estimated to

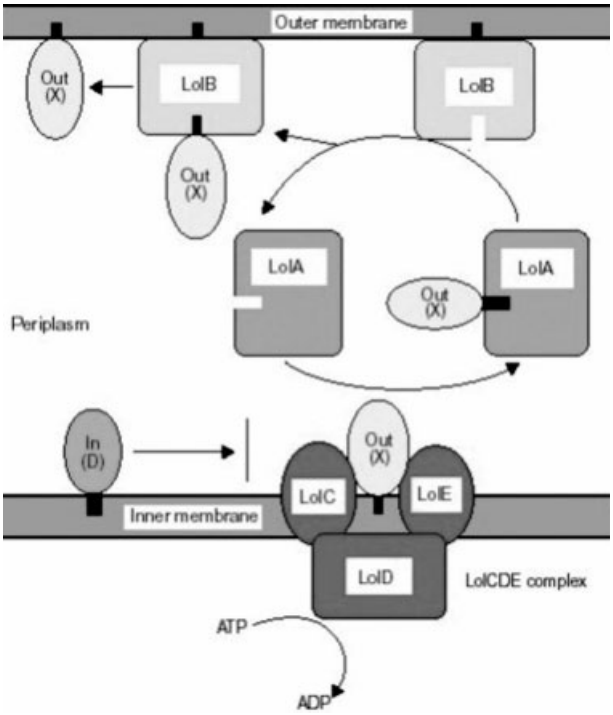


Fig. 8.6 Anchoring of lipoproteins into the outer membrane. The lipoprotein is detached from the inner membrane by the LolCDE complex, an energy-dependent process, bound by LolA, transferred to the outer membrane protein LolB which facilitates anchoring in the outer membrane. T. Yakushi, et al. **2000**, *Nat. Cell Biol.* 2, 212–218; Fig. 7.

be 150–300. Therefore, to complete the outer membrane localization, Lpp must be transferred from LolB to the outer membrane. It has been speculated that both LolA and LolB function catalytically to cycle the localization of the outer membrane-directed Lpp. While the protein moiety of LolB seems to be exposed to the periplasm, allowing interaction with the Lpp-LolA complex, the lipid moiety anchors LolB to the outer membrane. LolA is required not only for the release of Lpp, but also for Pal, NlpB and LolB and therefore seems to be involved in the outer membrane localization of most if not all lipoproteins in *E. coli*. Since both *lolA* and *lolB* are present in most if not all Gram-negative bacteria, this observation indicates that both genes are generally involved in the localization of outer membrane lipoproteins in Gram-negative bacteria.

8.6
Protein Translocation Pathways Through Both Membranes of Gram-negative Bacteria

Protein secretion into the extracellular milieu is required for numerous aspects of the bacterial life cycle, including nutrient acquisition and virulence-factor expression. In Gram-negative bacteria, proteins must cross the periplasm and the outer membrane in addition to the cytoplasmic membrane. Secretion across the outer membrane represents a different problem from translocation of unfolded proteins across the inner membrane. Proteins acquire structure in the periplasm, including the formation of disulfide bonds (see Section 7.3.2), and may completely fold before crossing the outer membrane. In addition, ATP and other sources of energy are absent from the outer membrane. Secretion systems must therefore be self-energized or have mechanisms for harnessing the energy available at the inner membrane. Gram-negative bacteria have evolved six different pathways for protein secretion: (1–5) the type I through V secretion pathways and (6) vesicle-mediated export. These pathways can be classified into two groups, according to their dependence upon the Sec pathway. Proteins that follow the Sec pathway have a signal peptide and are secreted in two steps. They are translocated to the periplasm via the Sec machinery, then cross the outer membrane and are transported to the extracellular medium either by a complex multicomponent system or by vesicles. Extracellular proteins of type II, IV and V are secreted by the two-step mechanism. Proteins using the other pathway do not have a signal peptide and are exported from the cytoplasm to the extracellular medium in one step, with no periplasmic intermediate. The type I and type III pathways are one-step mechanisms by which the secreted proteins cross directly from the cytoplasm to the bacterial surface.

8.6.1
The Type I Secretion Pathway

Many pathogenic bacterial species secrete degradative enzymes and pore-forming toxins to counteract host defenses. Several of these proteins are transported across the two membranes, using ATP-binding cassettes containing transporters (also called ABC exporters), and this secretion pathway has been denoted as the type I secretion pathway. It is exemplified by the HlyA system present in some pathogenic *E. coli* strains. The *hlyA* gene codes for a hemolysin which is secreted into the extracellular milieu where it attacks erythrocytes. The HlyA hemolysin is a lipid-modified polypeptide containing a domain with 11–17 nine-amino-acid repeats (LxGGxGND). These repeat domains bind calcium and interact with host cells, triggering HlyA insertion into the plasma membrane and leakage of the cytoplasmic content of target cells, such as erythrocytes. Similar repeats have been identified in other secreted proteins; and this group of polypeptides is referred to as the *family of repeat toxins*.

After synthesis in the bacterial cytoplasm, HlyA is modified by N-acylation at two lysine residues with myristate or palmitate in a reaction that requires the pro-

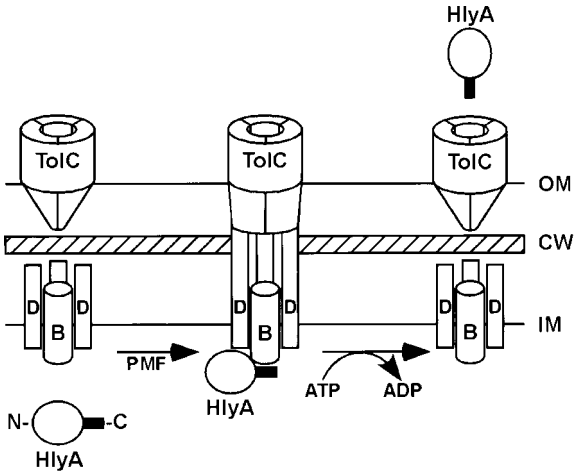


Fig. 8.7 Secretion of hemolysin (HlyA) occurs through type I secretion. After synthesis, HlyA interacts with its cognate ABC transporter complex (one subunit of ATP-binding cassette HlyB and three subunits of HlyD) in a proton-motif force (PMF)-dependent manner. HlyA is translocated in one single step across both membranes in a reaction that requires ATP hydrolysis and interaction of HlyD with the trimeric TolC. V.T. Lee, O. Schneewind **2001**, *Genes Dev.* 15, 1725–1752; Fig. 5.

duct of *hlyC*, the acyl carrier protein (ACP) and ATP (Fig. 8.7). The HlyC acyltransferase acylates the ε-amino groups of lysine within HlyA, using thioester-linked fatty acids (ACP) as substrate. Next, the acylated hemolysin is secreted across both membranes in a single step; and this process needs the three proteins HlyD, HlyB and TolC (see Fig. 8.9), the PMF and ATP. HlyD assembles into a trimer, spans the inner and outer membrane and binds directly to both TolC and HlyB (Fig. 8.7). Secretion of HlyA occurs in several steps: (a) formation of a complex between HlyA and HlyB/HlyD, (b) binding of ATP by HlyA/$HlyB_n$/$HlyD_3$, (c) conformational change of HlyD resulting in the formation of a HlyA/$HlyB_n$/$HlyD_3$/$TolC_3$ complex, (d) HlyB-mediated movement of HlyA across the inner membrane and ATP hydrolysis and (e) movement of HlyA across the outer membrane. The first four steps are fueled by alterations in the folding state of HlyB and are triggered by the binding, hydrolysis and release of ATP. In addition, the PMF across the inner membrane is needed and reagents that disrupt the proton gradient abolish the initial stages of HlyA transport.

8.6.2
The Type II Secretion Pathway

The type II secretion pathway is encoded by 12–15 genes and specifically supports the transport of a group of seemingly unrelated proteins across the outer membrane, including bacterial toxins and type IV pili. Before these proteins enter the type II secretion pathway, they have to first translocate across the cytoplasmic

membrane via the Sec system and then fold into a secretion-competent conformation in the periplasm. Proteins secreted by the type II pathway include proteases, cellulases, pectinases, elastases, phospholipases and toxins. Expression of the genes coding for these proteins and, in some cases, the secretion genes themselves are under quorum-sensing control (see Section 6.2.1.1) or are strictly regulated by the environment at the side of colonization. This strict regulation ensures that virulence factors are secreted only when the bacteria have reached their correct location and obtained a critical mass. Species identified to date as harboring type II secretion systems are members of the proteobacteria and most of them appear to be extracellular pathogens.

The proteins involved in secretion across the outer membrane have been assigned a unified Gsp (for *g*eneral *s*ecretory *p*athway) nomenclature. Twelve Gsp proteins (GspBCFGHIJKLMNO) are positioned in the inner membrane, two in the outer membrane (GspDS) and one (GspE) resides in the cytoplasm (Fig. 8.8). The topology of the 11 inner membrane proteins has been deduced from analysis of fusions to alkaline phosphatase, β-lactamase or β-galactosidase. GspF and GspO are polytopic, spanning the membrane eight and three times, respectively. Both have appreciable segments in the periplasm and one large cytoplasmic domain. The remaining nine proteins span the membrane only once, with their N-termini exposed to the cytoplasm. Four of these are synthesized as precursors with a prepilin signal sequence (GspGHIJ). GspE associates with the cytoplasmic membrane when GspL is present, but remains cytoplasmic when produced alone. One component, GspD or *secretin*, is inserted into the outer membrane and oligomerizes into a dodecameric ring structure with an inner diameter of 7.6 nm. It is assumed that polypeptides move through the lumen of GspD. Some GspD proteins require the outer membrane chaperone GspS for proper folding and activity *in vivo*, such as that of *K. oxytoca*. Most of the proteins that constitute the type III secretion system are integral membrane proteins. What makes the role of so many proteins puzzling is the observation that translocation of substrate proteins to the periplasm occurs in the complete absence of the accessory proteins. The idea is that, when the intact complex is present, secretion substrates are not freely diffusible in the periplasm, but may associate with the complex. The specific function of individual components remains an open question, especially because many serve exclusively structural roles.

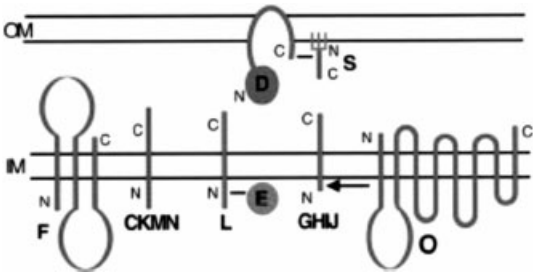


Fig. 8.8 Arrangement of the type II secretion proteins in the cell envelope. While the proteins CFGHIJKLMNO are bi- or polytopic inner membrane proteins, protein D and lipoprotein S are anchored in the outer membrane. M. Russel **1998**, *J. Mol. Biol.* 279, 485; Fig. 2.

Besides being involved in the secretion of bacterial toxins, several Gsps exert specific functions during the assembly of type IV pili which are produced by *P. aeruginosa, N. gonorrhoeae* and other Gram-negative pathogens. These pili are thought to retract in a coordinated fashion. Pilus adhesion to an immobilized surface and pilus retraction can pull bacteria such as *Myxobacteria* or *Neisseria* in one direction. Type IV pili also occur in *P. aeruginosa* where they are involved in adhesion to respiratory epithelia and are composed of the major pilin subunit PilA. PilA is synthesized as a precursor with a signal anchor sequence and a unique N-terminal prepilin signal sequence. After translocation of the precursor by the Sec machinery through the inner membrane, the prepilin signal sequence is cleaved and the liberated amino-group is N-methylated by GspO using *S*-adenosyl-methionine as a cofactor (Fig. 8.9). Mature translocated PilA is assumed to polymerize at the inner membrane; and the assembled pilus may be extruded through the central cavity of the outer membrane secretin.

All the exoproteins use the Sec pathway for translocation through the inner membrane. But in *P. aeruginosa*, it could be shown that two different phospholipases make use of the Tat pathway to cross the inner membrane. Since this pathway transports folded proteins, it is an open question at the moment why these two enzymes have to be translocated in their folded form.

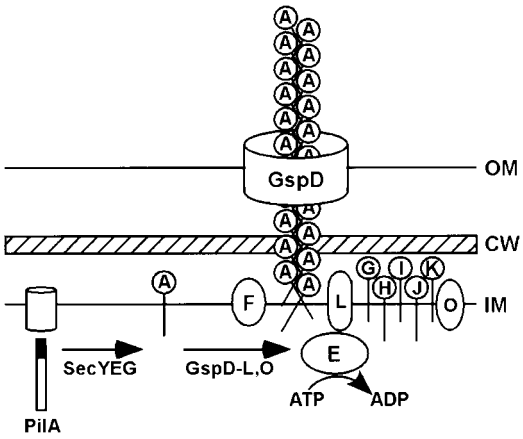


Fig. 8.9 Assembly of type IV pili. In *P. aeruginosa*, the type IV pilin subunits are synthesized with a signal sequence. After translocation via the Sec system, removal of the signal peptide by signal peptidase I, the N-terminal amino group is methylated by the GspO protein which uses *S*-adenosyl methionine as cofactor. Pilus polymerization is dependent on the cytoplasmic GspE and the inner membrane Gsp proteins GspFGHIJKLO followed by extrusion through the central cavity of the outer membrane secretion (GspD). V.T. Lee, O. Schneewind **2001**, *Genes Dev.* 15, 1725–1752; Fig. 7.

The Curli Assembly Machinery

At least five proteins are involved in assembling curli fibers on the cell surface of *E. coli* cells. These proteins are encoded by the divergons *csgBA* and *csgDEFG* (Fig. 8.10). The type II secretion machinery is responsible for translocating the curli subunits across the inner membrane, while transport across the outer membrane remains elusive. CsgG is a 30-kDa lipoprotein that localizes to the periplasmic face of the outer membrane. Two alternative functions have been suggested for CsgG: it may bind to curli subunits to stabilize them in the periplasm or it might form pores through which the subunits escape the periplasm. CsgF might be required for full CsgB activity and could modify CsgA to make it polymerization-competent, acting as a molecular chaperone for CsgA. The role of CsgE is less clear. It could either transfer the CsgA and CsgB proteins from the CsgF chaperone to the CsgG pore as shown (Fig. 8.10) or, alternatively, it may contribute to the stability of the two curli fiber proteins or both. The function of CsgD is completely elusive.

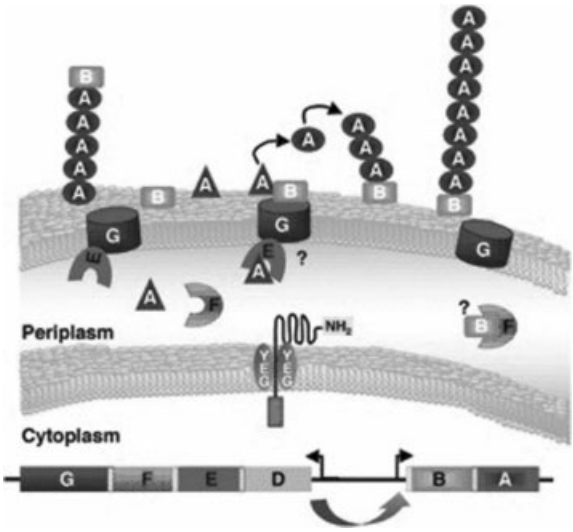


Fig. 8.10 Current model of curli fiber regulation and assembly. All curli subunits except CsgD start with a signal sequence required for translocation into the periplasm by the Sec pathway. CsgA and CsgB are the major and minor curli subunits, respectively, kept in a polymerization-competent form by the CsgF chaperone. CsgG might form a pore in the outer membrane through which CsgB and CsgA escape where CsgB acts as a nucleator. CsgE could transfer CsgA and CsgB to CsgG as indicated. M.R. Chapman, et al. **2003**, *ASM News* 69, 121–126; Fig. 2. (This figure also appears with the color plates.)

8.6.3

The Type III Secretion Pathway

The type III secretion (TTS) pathway is used by pathogenic and symbiotic Gram-negative bacteria to deliver *effector proteins* within the membrane or even the cyto-

sol of eukaryotic cells with which these bacteria interact. These effector proteins bring about profound changes in the cytoskeletal dynamics, signal transduction and cell–cell communications. The effectors injected by some systems paralyze phagocytes and allow bacteria to proliferate extracellularly, whereas effectors injected by other systems stimulate epithelial cells to phagocytose bacteria and allow them to spread intracellularly. These secretion machines are essential for the pathogenicity of *B. bronchiseptica*, *B. pseudomallei*, *C. psittacii*, *E. amylovora*, *E. crysanthemie*, pathogenic *E. coli*, *P. aeruginosa*, *P. syringae*, *R. solanacearum*, *Rhizobia* species, *S. typhimurium*, *S. flexneri* and *X. campestris* as well as three *Yersinia* species. TTS systems are complex secretion and delivery machines and need more than 25 genes. In animal pathogens, these secretion machines deliver anti-host virulence determinants into mammmalian cells, while they cause disease in susceptiple plants and trigger the hypersensitive response in resistant plants. Genetic regulation of the TTS systems occurs in at least two distinct steps: (a) expression of genes required for assembly of the secretion apparatus and (b) expression of genes whose products are substrates for this secretion system. The binding of bacterial and host cell surface receptors appears to be one mechanism that activates type III machines therefore also called the contact-dependent secretion system. General features of this system include:

1. The secreted proteins do not have a typical N-terminal sequence characteristic of proteins exported via the Sec pathway.
2. The export machinery directs the translocation of the target proteins through two membranes without cleavage of their N-termini.
3. An inducing extracellular signal, usually resulting from the interaction with host cells, is required for complete activation of the secretory apparatus.

Salmonella species employ the type III machine for entry into epithelial cells and inject factors that activate and stabilize the polymerization of actin filaments in the host cytoplasm. A second type III pathway is used for *Salmonella* survival within macrophages. A *Yersinia* type III secretion system prevents the phagocytic killing of microbes that have adopted an extracellular life style. The same secretion system allows enteropathogenic *E. coli* to damage epithelial cells and provides a microbial growth advantage in the human intestines. Since several type III genes display striking homology to basal body genes, specifiying a machinery that is dedicated to the assembly of flagella, we will first focus on this system.

Assembly of the Bacterial Flagellum

The bacterial flagellum is composed of three parts: the *basal body* (consists of three rings and a connecting rod), the *hook* and the *flagellar filament*. The basal body consists of three rings in Gram-negative bacteria (MS, P, L) and the rod. Basal body assembly starts with the formation of the MS ring (composed of FliF) within the inner membrane (Fig. 8.11) and is required for the assembly of the rod (FlgBCFG) and the P (FlhI) and L (FlgH) rings. In a second phase, the hook

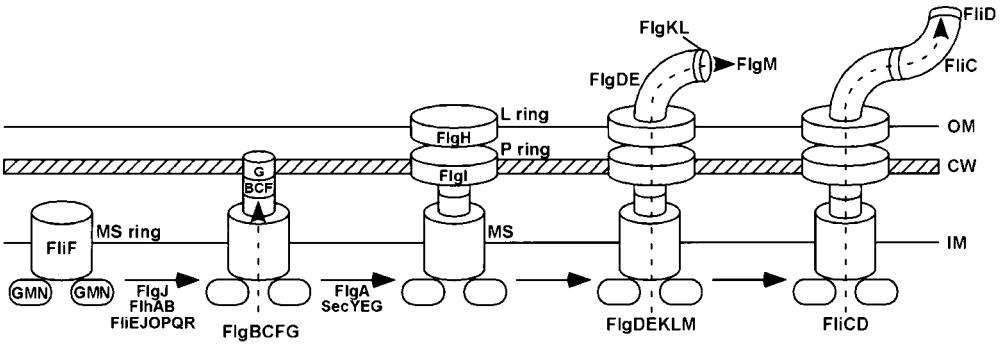


Fig. 8.11 Assembly of the flagellum. First, the FliF protein is inserted into the inner membrane forming the MS ring. Addition of further Fli and Flg proteins complete the basal body and anchor it into the cell envelope including the P and L ring. Next, the hook is formed (FlgDE proteins), sealed by the capping proteins FlgKL. FliD is a dual function protein which forms the flagellar cap and is the assembly site for the flagellin protein FliC. V.T. Lee, O. Schneewind **2001**, *Genes Dev.* 15, 1725–1752; Fig. 9.

(FlgDEKLM) is assembled and then triggers polymerization of the filament. The filament consists of 11 flagellin subunits (FliC) in two complete right-handed helical upward turns of the filament and is sealed by a pentamer of the capping protein (FliD). The flagellin filament protrudes 3–12 μm above the bacterial surface. What is the energy source that transports the subunits to the tip? Two models are envisaged. In the first model, subunit proteins could line the interior of the flagella filament from the base to the tip. The addition of another subunit to the base would push all subunits further into the filament and cause the most distal subunit to bind the structure at its tip. In the second model, subunits are transported from the base to the tip by utilizing the rotational force of the flagellum. Rotation of the flagellar filament propels bacteria in liquid media and is fueled by the PMF as well as ATP hydrolysis. Two important principles govern flagellar assembly:

1. The complex structure is built by adding the different subunits at the tip (contrary to pili). The basal body rod, the hook and the filament are believed to form a channel for the transport of subunits to the tip of the structure.
2. The completion of rings, rods or hook serves as checkpoints to promote expression and transport of new subunits for subsequent parts of the structure.

The genetic information encoding the components and targets of the TTS machinery in *Salmonella* is located in a contiguous region of the chromosome, which constitutes an example of a *pathogenicity island*. Other type III systems are also known to be encoded in either discrete chromosomal regions or in virulence plasmids. These observations suggest that these genetic determinants have their origins in a common anchestor.

TTS systems consist of the following elements:

- Approximately 15 proteins are required to construct the TTS apparatus, the *injectisome*, which spans the bacterial envelope and consists of a transmembrane basal body and an external component, a needle-like structure instead of a hook and a filament.
- Two translocators insert into the outer membrane and form a pore through which effectors transit.
- Effectors are injected into eukaryotic cells and interfere with cellular processes as described above.
- Specific molecular chaperones associate in the bacterial cytoplasm with a subset of proteins secreted by the TTS system (see Section 7.2.2).
- Specific transcriptional regulators.

The injectisome of *Y. pestis* and *Y. enterocolitica* consists of 27 *Y*op *s*ecretion (Ysc) proteins and appears as two pairs of rings that are anchored to the inner and outer membranes of the bacterial envelope, joined by a central rod and supporting a hollow needle about 10 nm thick and 60 nm long. It is assumed that the injectisome serves as a hollow conduit through which the secreted proteins travel across the two bacterial membranes and the peptidoglycan in one step. How is the length of the needle determined? It has been suggested that YscP controls the length of the needle by acting as a molecular ruler during the stepwise assembly of the injectisome. The two ends of YscP serve as anchors. One end is attached to the basal body, whereas the other is connected to the growing tip of the needle. When the needle reaches its mature length, YscP is fully stretched and signals, via its internal anchor, to the secretion apparatus to stop further export of YscF, the needle subunit.

8.6.4
The Type IV Secretion Pathway

Type IV secretion systems are found in plant and animal pathogens as well as in symbiotic bacteria. These systems are known to transport proteins or nucleoprotein complexes and are commonly associated with virulence. Native substrates that depend upon type IV secretion systems for their transfer are the pertussis toxin in *B. pertussis*, the CagA protein in *H. pylori* (forms cylinders that might induce host cytoskeletal rearrangements) and the T-DNA of *A. tumefaciens*. These proteins are required for virulence in each of these organisms. The proteins comprising type IV secretion systems are anchestrally related to proteins of the mating pair formation systems required for plasmid conjugal transfer. Both systems are thought to form a pilus required for substrate transfer to recipient cells, although this structure has only been confirmed in *Agrobacterium* and *E. coli*. *A. tumefaciens* contains two type IV systems encoded by the *virB* and *trb* operos of pTiC58; and the two systems have different specificities and mediate the transfer of the Ti plasmid or T-DNA, respectively. Secretion of *B. pertussis* holotoxin by the Ptl system and transfer of the oncogenic T-DNA by the VirB transporter will be discussed.

Secretion of *B. pertussis* Holotoxin by the Ptl System

The Ptl system of *B. pertussis*, the causative agent of the pertussis disease (whooping cough), is responsible for the transport of just one single virulence factor, the pertussis toxin. By ADP ribosylation, this toxin activates the two eukaryotic proteins G_1 and G_0 that are involved in signal transduction. The active pertussis toxin consists of the enzymatically active S1 subunit that interacts with the B oligomer formed by one copy each of S2, S3 and S5 and two copies of S4 (Fig. 8.12). Upon arrival in the periplasm, the S1 subunit associates with the outer membrane and serves as a nucleation site for the assembled B oligomer. Secretion of the holotoxin through the outer membrane needs the Ptl system that consists of nine different proteins assumed to form a channel in the outer membrane.

Transfer of Oncogenic T-DNA Into Plant Cells

The best-studied type IV secretion pathway is the VirB system of *A. tumefaciens*. The VirB protein forms a long thin filament, the T pilus, 10 nm long and about 3.8 nm in diameter. At least ten of the 11 VirB proteins (VirB2–VirB11) associate to form the pilus structure, spanning from the cytoplasm through the inner membrane, the periplasmic space and the outer membrane to the outside of the cell (Fig. 8.13). VirB4 and VirB11 are both homodimeric integral membrane proteins possessing ATPase activity and may energize the transport. Proteins VirB6, VirB8 and VirB10 form the transcytoplasmic pore and link the cytoplasmic and outer

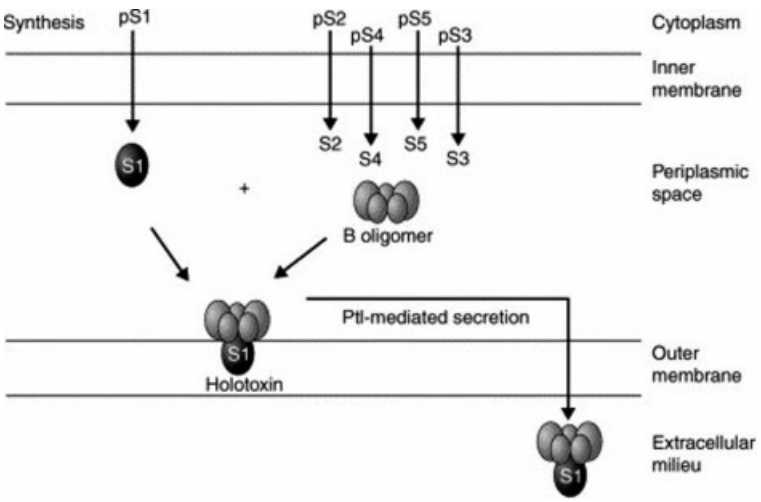


Fig. 8.12 Model for secretion of pertussis holotoxin via the type IV secretion pathway. The five different subunits are independently translocated through the inner membrane via the Sec system. The S1 subunit associates with the outer membrane and serves as a nucleation site for the B oligomer. In the last step, the pertussis holoenzyme is secreted through the outer membrane by the Plt system. The Plt system consists of nine proteins which most probably forms a channel to allow secretion of the holotoxin. D.L. Burns **2003**, *Curr. Opin. Microbiol.* 6, 1–6; Fig. 3.

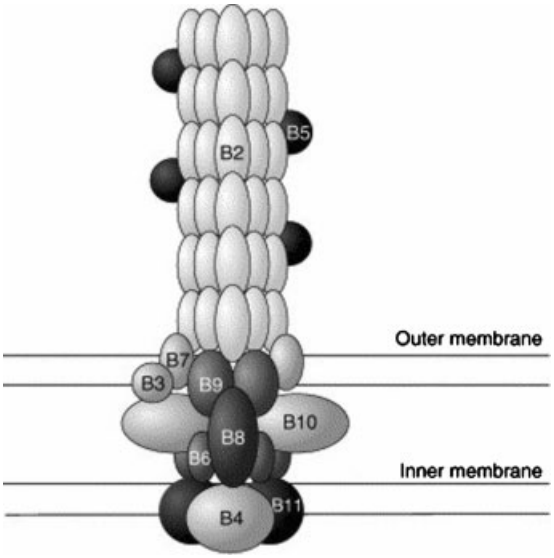


Fig. 8.13 Schematic drawing of the VirB transporter of *A. tumefaciens*. The VirB transporter is composed of the proteins VirB2 through VirB11, and the localization of each protein within this large complex is shown. The pilus rod through which the T-DNA is transferred from the bacterial cell into the plant cell consists of VirB2 with some attached molecules of VirB5. D.L. Burns **2003**, *Curr. Opin. Microbiol.* 6, 1–6; Fig. 2.

membrane VirB subcomplexes. VirB3, VirB7 and VirB9 are found at or in the outer membrane, where VirB9 by virtue of a β-pleated sheet structure in its C-terminus comprises all or part of an outer membrane pore structure. VirB2, the major subunit of the T-pilus, is processed from a 12.5-kDa proprotein to the 7.0-kDa mature peptide. Following processing, the peptide undergoes a cyclization reaction such that the N- and C-termini are joined via an intracellular covalent head-to-tail peptide bond (for additional examples of cyclic proteins, see Section 6.5.5). The VirB5 subunits are associated with VirB2 and their function is unknown. The last VirB protein, VirB1, is not part of the transporter, has a motif found among lytic transglycosylases and might cause local lysis of the peptidoglycan during transporter assembly.

In general, the T-pilus is essential for T-DNA transfer and virulence. In particular, the T-pilus could serve as a conduit for the following components: (a) the T-pilus subunits are transferred through and out onto the growing tip of the pilus, (b) the single-stranded T-DNA has its 5′ end covalently linked to the VirD2 protein and (c) the VirE2 protein and VirF proteins are exported into plant cells.

8.6.5
Type V Secretion Pathway (Autotransporter)

Autotransporters are extracellular proteins which catalyze their own secretion. One example is the IgA protease secreted by *N. gonorrhoeae*, able to cleave anti-

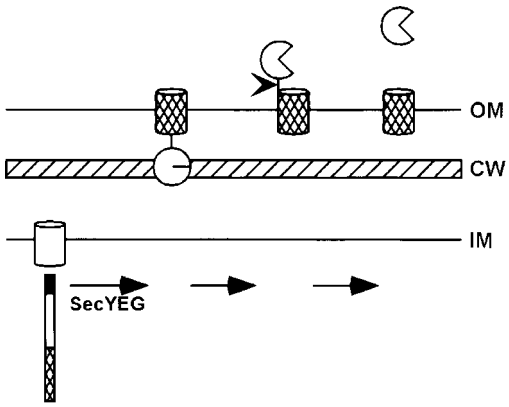


Fig. 8.14 Autotransporters. Proteins are synthesized as pre-proenzymes and translocated across the inner membrane via the Sec system. Next, the C-terminal domain of the proprotein inserts into the outer membrane and translocates its N-terminal domain through its lumen. In the last step, the protease activity of the N-terminal domain cleaves within the flexible linker connecting the two domains (indicated by the arrowhead) which leads to its liberation. V.T. Lee, O. Schneewind **2001**, *Genes Dev.* 15, 1725–1752; Fig. 4.

bodies on mucosal surfaces. This enzyme is synthesized as a preproenzyme and translocated through the Sec pathway into the periplasm while its signal peptide is cleaved off (Fig. 8.14). The resulting proenzyme is presumed to be only partially folded, and its C-terminal β-domain assumes a β-barrel structure that inserts into the outer membrane and functions as an autotransporterautotransporter",4> for the N-terminal domain. When the N-terminal protease domain is exposed on the bacterial surface, it cleaves the proenzyme at the junction between the N- and C-terminal domain, releasing the proenzyme from the bacterial surface into the extracellular milieu. The proprotease matures into the 106-kDa IgA protease and the small stable α-protein. The biological function of the IgA protease is to cleave antibodies on mucosal surfaces involved in complement-mediated killing of bacteria.

8.6.6 Vesicle-mediated Export

Gram-negative bacteria are able to produce membrane vesicles which are derived from the outer membrane. These *o*uter *m*embrane *v*esicles (OMVs; 50–200 nM in diameter) are constantly being discharged from the surface of the cell during bacterial growth and include outer membrane proteins, LPS and phospholipids. It could be shown that OMVs have an important biological function in the secretion and delivery of bacterial toxins to mammalian cells, indicating a mechanism of vesicle trafficking that involves the transfer of protein cargo from prokaryotic to eukaryotic cells. One proven example is cytotoxic ClyA protein, a pore-forming toxin

of *E. coli*. This protein is synthesized with a signal peptide and translocated into the periplasm via the Sec pathway. Within the periplasm, ClyA undergoes an oligomerization and forms ring-like pores in the outer membrane. Toxicity assays showed that ClyA present in OMVs is at least eight times more toxic than purified ClyA. Another example is the heat-*l*abile entero*t*oxin (LT) which is secreted from *e*nero*t*oxigenic *E. coli* (ETEC) via vesicles. These ETEC vesicles serve as specifically targeted transport vehicles that mediate entry of active enterotoxin and other bacterial envelope components into host cells and demonstrate a role in virulence. OMVs seem to contain a subset of periplasmic proteins, suggesting that *E. coli* could have a crude protein-sorting mechanism that selects which proteins are exported in OMVs. If this turns out to be true, it will be interesting to elucidate how the proteins and cargo in the periplasm are sorted, how sorting is regulated and whether this process can be used as a tool for delivery of proteins into eukaryotic cells.

There remain many open questions. Is OMV formation regulated and are vesicles being formed *in vivo* on infection of animals? How are proteins and cargo in the periplasm sorted to OMVs? Is the lipid content of the vesicles unique? How essential is vesicle formation to the intracellular pathogenesis of Gram-negative bacteria?

8.7 Cell Wall-anchored Surface Proteins

Most Gram-positive bacteria code for enzymes called sortases which are able to covalently anchor surface proteins to the cell wall peptide cross-bridges. This process needs four different components:

1. The proteins to be anchored carry a signal sequence at their N terminus ensuring translocation through the inner membrane via the Sec system.
2. In addition, these proteins contain an about 35-amino-acid residues tripartite cell wall sorting signal near their C terminus. This signal consists of a sorting motif followed by a hydrophobic domain and a tail of mostly positively charged amino acid residues.
3. A cytoplasmic membrane-anchored sortase with its transpeptidase activity exposed on the outside of the membrane. In general, two different sortases have been identified in Gram-positive bacteria designated as sortase A and B. But there are species coding for up to six potential sortases.
4. The sortase recognizes the sorting motif and the cell wall cross-bridges. They anchor the target proteins on the lipid II molecules before they become inserted into the cell wall.

Anchoring of surface proteins to the cell wall has been detected and studied extensively in *S. aureus*. Here, two sortases have been identified, where sortase A recognizes the sorting motif LPXTG (where X stands for any amino acid) and sortase B

the motif NPQTN. The first identified protein for sortase A is protein A. The N-terminal part of this protein carries five Ig-binding domains which are able to bind to the Fc portion of antibodies and cause precipitation of immunoglobulins. After translocation through the inner membrane and removal of the signal peptide, protein A is bound by sortase A as shown in Fig. 8.15. This transpepidase cleaves protein A at its sorting motif between the threonine and the glycine, resulting in the formation of a thioester enzyme intermediate. Next, a nucleophilic attack of the free amino group of lipid II at the thioester bond leads to the transfer of protein A to lipid II forming an amide bond between the surface protein and the pentaglycine cross-bridge. Incorporation of protein A into the murein occcurs in two steps. First, the lipid-linked protein is covalently linked to peptidogylcan via the transglycosylation reaction, followed by cross-linking between the murein pentapeptide subunit and other cell wall peptides.

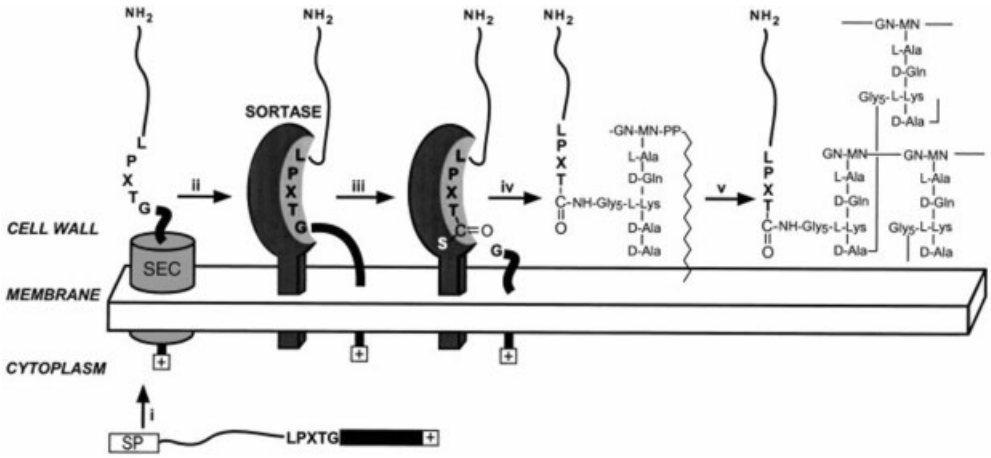


Fig. 8.15 Anchoring of proteins on the cell wall of *S. aureus*. The surface proteins to be anchored contain two different signals, a signal sequence at the N-terminus and a sorting signal close to the C-terminus. After translocation through the inner membrane via the Sec system and removal of the signal sequence (I), the polypeptide is recognized by the sortase through the sorting sequence (LPXTG, where X stands for any amino acid), cleaved between the threonine and the glycine and first linked in a transpeptidase reaction to the enzyme through a thioester linkage (III). Next, a nucleophilic attack of the free amino group of lipid II releases the surface protein from the sortase and links it to the amino group of the pentaglycine (IV). In the last two enzymatic steps, the surface protein is transferred to the cell wall via the transglycosylation reaction followed by cross-linking to other cell wall peptides (V). S.K. Mazmanian, et al. **2001**, *Mol. Microbiol.* 40, 1049–1057; Fig. 1.

8.8
S-Layers

S-layer proteins have already been introduced (see Section 1.5.1). They surround cells of many different Eubacteria and Archaea; and we will briefly describe here what is known about how S-layer proteins reach their final destination.

Secretion of the S-protein has been studied extensively with *A. salmonicida*. The S-protein called VapA is first translocated through the inner membrane using its signal sequence and the Sec system (Fig. 8.16). Having arrived in the periplasm, it binds to the hypothetical S-protein-specific chaperone ApsE which escorts VapA to the outer membrane protein SpsP. This protein exhibits homology with the group of secretins and is therefore believed to form a pore in the outer membrane through which the S-layer protein VapA is secreted. The further steps which lead to integration of VapA into the S-layer and binding of the LPS have not been elucidated. Not all S-layer proteins contain a signal sequence; and these are secreted using the type I secretion pathway (see above).

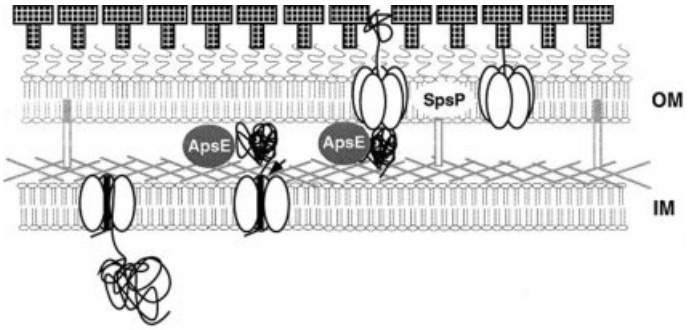


Fig. 8.16 Suggested pathway for the secretion of the S-layer protein VapA of *A. salmonicida*. The VapA protein is first translocated through the inner membrane using the Sec system. After arrival in the periplasm, VapA is bound by the putative ApsE chaperone preventing aggregation or/and complete folding. VapA is transferred to the SpsP complex acting as a secretin to allow secretion into the extracellular milieu. The further steps of insertion into the S-layer and binding to the LPS are unknown. L.A. Ferández, J. Berenguer **2000**, *FEMS Microbiol. Rev.* 24, 21–44; Fig. 9.

Suggested Reading

V. T. Lee, O. Schneewind **2001**, Protein secretion and the pathogenesis of bacterial infections, *Genes Dev*. 15, 1725–1752.

H. Mori, K. Ito **2001**, The sec protein-translocation pathway, *Trends Microbiol.* 9, 494–500.

H. Tjalsma, et al. **2004**, *Microbiol. Mol. Biol. Rev*. 68, 207–233.

H. Ton-That, L. A. Marraffini, O. Schneewind **2004**, *Biochim. Biophys. Acta* 1694, 269–278.

K. H. M. van Wely, J. Swaving, R. Freudl, A. J. M. Driessen **2001**, Translocation of proteins across the cell envelope of Gram-positive bacteria, *FEMS Microbiol. Rev*. 25, 437–454.

9
Stress Genes and Their Regulation

Bacteria are frequently exposed to changes in their environmental conditions. Important stress factors are: a sudden increase or decrease in temperature (called heat shock and cold shock, respectively), increase or decrease in the external pH value (alkaline and acid shock), generation of reactive oxygen species (oxidative stress), increase or decrease in the external osmolarity (hyper- and hypoosmotic stress) and limitations to nutrients (starvation stress). To overcome such environmental cues, bacteria have evolved genetic programs which allow a fast adaptation to the new situation, resulting in the transient increased expression of a subset of their genes called stress genes coding for stress proteins. Some stress genes respond to just one specific stress factor and these have been designated *specific stress genes* (e.g., the nine genes of the HrcA regulon of *B. subtilis*). In contrast, genes which are induced by several stress factors have been termed *general stress genes* (e.g., the σ^B regulon of *B. subtilis*) General stress responses provide cross-protection against a wide variety of environmental cues regardless of the initial stimulus.

9.1
The Stress Response Pathway

Studies carried out over the past 25 years, mainly with *E. coli* and *B. subtilis* as genetic model organisms, have revealed that adaptation to most stress factors follows a general stress response pathway presented in Fig. 9.1. The pathway starts with the *prestimulus state* where cells are in the exponential growth phase. Then, these cells are exposed to a physical or chemical stress factor, such as a sudden increase in the growth temperature, a heat shock. The stress factor creates a *signal* which is recognized by a *sensor* molecule, in most cases a protein. In a few cases, the sensor is covalently coupled to the gene to be expressed. If the stress factor acts on the outside of the cytoplasmic membrane, it has to be transmitted into the cytoplasm. This occurs by two well known mechanisms, the two-component signal transduction pathway (see Section 6.2.8.5) or the anti-sigma factor sequestering ECF sigma factors (see Section 6.2.1). The sensor either directly or indirectly interacts with a *transcriptional regulator* which in turn is responsible for the en-

Dynamics of the Bacterial Chromosome. Structure and Function. W. Schumann
Copyright © 2006 Wiley-VCH Verlag GmbH & Co. KGaA
ISBN 3-527-30496-7

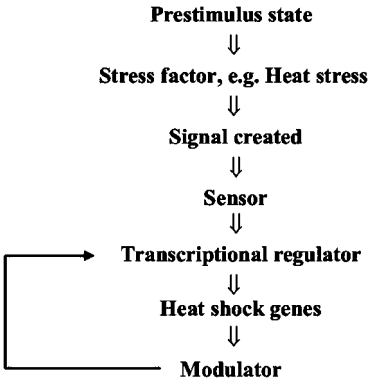


Fig. 9.1 The stress reponse pathway. See text for detailed explanation of the different steps of the pathway.

hanced expression of the appropriate stress genes (from a few up to about 200, depending on the stress signal). Since there are three different types of transcriptional regulators (alternative sigma factors, transcriptional activators or repressors), the stress response has to lead to the activation of the first two and the inactivation of the latter. Since in most cases the stress response is transient, the genetic program comprises a *modulator*, too, ensuring turn-off of the stress genes after adaptation, termed feedback inhibition.

9.2 The Heat Shock Response

The heat shock response is universally conserved in evolution and has evolved to detect and remove denatured and misfolded proteins, collectively termed non-native proteins, occurring especially after exposure of cells to sudden temperature increases. The heat shock response is characterized by the transiently increased synthesis of proteins called *h*eat *s*hock *p*roteins (HSPs) that are involved in protein folding, refolding of non-native proteins or their degradation. Abrupt increases in temperature occur, for example, when we eat raw vegetables and fruits. The bacteria living on this food are at a sudden exposed to 37 °C in our mouth. Another example are bacteria living in the soil. During a hot, sunny summer day, they may first reside in the shade of a tree. If the shade moves, these bacteria are exposed within seconds to the full sunlight, leading to an uprupt increase in temperature. A third example are pathogenic bacteria of mammals which live either within their host or outside. While the temperature within their host is always 37 °C, the outside temperature is normally below this value. This group of bacteria explores the 37 °C temperature to induce expression of their virulence genes.

A major problem associated with such a heat shock is partial unfolding and misfolding of proteins, leading to the exposure of hydrophobic amino acid residues on the outside which are normally buried in the interior of the polypeptide chains, which makes these proteins ''sticky''. These sticky proteins are prone to aggregation; and large aggregates cannot be dissolved by the protein control me-

chanisms within the cell. Therefore, all cells try to prevent the formation of such aggregates by two different defense strategies: (1) molecular chaperones and (2) ATP-dependent proteases. Important HSPs are molecular chaperones and proteases. They are essential for normal growth and their activities become particularly important under those stress conditions which lead to the accumulation of unfolded or incorrectly folded proteins. The regulatory mechanisms for HSP induction differ widely between eubacteria and involve three general mechanisms: alternative RNA secondary structures, transcriptional repressors and alternative sigma factors.

Detection of the Heat Shock Response

Detection of the heat shock response is based on a mistake. In the early 1960s, F. Ritossa, an Italien scientist, was studying the puffing pattern of the giant chromosomes prepared from *Drosophila* larvae, where puffs represent transcribing genes. One morning, he discovered a puffing pattern which was quite different from what he had observed so far. Upon inquiry of his collegues, he learned that one of them inadvertently increased the temperature of the incubator used to keep the larvae at a constant temperature from 25 °C to 32 °C. Ritossa repeated this experiment several times with comparable results and concluded that a sudden increase in the growth temperature resulted in an altered gene expression pattern. Twelve years later (1974), A. Tissières, a Swiss scientist, separated radioactively labeled *Drosophila* proteins prepared from different tissues using SDS-PAGE and could show by this technique that indeed a novel proteins pattern occurs after a heat shock. He suggested calling these proteins *heat shock proteins*. In the late 1970s, heat shock proteins were also detected in *E. coli*; and we know now that they occur in all organisms, from bacteria to man.

Function of Important Heat Shock Proteins

What is the function of these heat shock proteins? The most prominent ones belong to one of two groups: molecular chaperones or ATP-dependent proteases. The function of both classes of proteins have been described in detail (see Section 7.2 and 7.4). Additional heat shock proteins have been identified, but their function during adaptation to a sudden temperature increase mostly remains elusive.

Bacterial Heat Sensors

Two different classes of heat sensors have been described, direct and indirect sensors. Direct heat sensors are either RNA molecules or repressor proteins able to adopt two alternative conformations dictated by the temperature, while indirect heat sensors are molecular chaperones or proteases which modulate the activity of transcription factors in response to non-native proteins. It should be mentioned that direct heat sensors do not allow a shut-off of the heat shock response at the high temperature. They need the low temperature to adopt their active conformation.

Bacterial Direct Heat Sensors

In the case of sensor RNA, several different transcripts have been described able to sense temperature. We will discuss the *rpoH* (*R*NA *p*olymerase *h*eat) mRNA present in *E. coli* and other bacterial species, including the Enterobacteriaceae, encoding the alternative sigma factor σ^{32}, and a transcript present in the Rhizobiae and coding for a small heat shock protein. Both types of transcripts contain long 5′ untranslated regions able to form secondary structures at low temperatures (around 30 °C) where the Shine–Dalgarno (SD) sequence is part of a base-paired region (Fig. 9.2). Under these conditions, ribosomes have difficulties recognizing and binding to the SD sequence. If the temperature increases, the secondary structure becomes unstable, depending on the absolute temperature, allowing the ribosomes to access the SD sequence (Fig. 9.2). The higher the temperature in the growth medium, the more unstable the secondary structure and the more frequent translation of the transcript occurs. Both RNA thermosensors regulate the expression of just one gene, but with different outcomes. While the rhizobial mRNAs code for a small heat shock protein involved in the binding of non-native proteins (see Section 7.2.2), the *rpoH* transcript codes for the heat shock sigma factor σ^{32} which in turn regulates expression of about 60 different genes in *E. coli*. The amount of active σ^{32} is not only regulated by translation, but in addition by stability (see below).

In the case of bacterial repressor proteins, we will discuss the RheA and TlpA transcriptional regulators. The RheA repressor has been discovered in *S. albus*, where is controls expression of the *hsp18* gene coding for a small HSP and its own gene, both forming a divergon. At low temperature, RheA is present in its active conformation; and it represses expression of *hsp18* and autoregulates its own expression. After heat challenge, RheA changes its conformation and dissociates from its operator, allowing a high level of *hsp18* transcription. The constitutive synthesis of HSP18 for prolonged periods at high temperature is also consistent with the *in vitro* findings that RheA is intrinsically temperature-sensitive.

The second example is the dimeric TlpA repressor identified as a virulence-encoded protein in pathogenic *S. typhimurium* strains. The TlpA protein consists of an N-terminal DNA-binding region and a long coiled-coil domain. As described

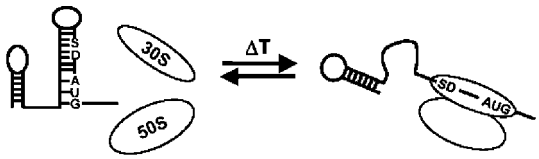


Fig. 9.2 RNA thermometers adopt alternative secondary structures dictated by the temperature. At low temperatures, the Shine–Dalgarno sequence and sometimes the start codon, as shown here are part of the stem of a secondary structure, making it difficult for the ribosomes to bind. At high temperatures indicated by the DT, the stem structure becomes unstable, now allowing the ribosomes to bind and initiate transcription of the downstream gene. F. Narberhaus **2005** *FEMS Microbiol. Rev.* 30, 3; Fig. 1b.

for RheA, TlpA is an autoregulatory repressor which uses its folding equilibrium to regulate the DNA binding activity. The mechanism is based on the long coiled-coil domain that spans more than two-thirds of the primary structure. In the TlpA repressor, this equilibrium is adjusted in such a way that it can act as a temperature sensor. Temperature upshifts lead to a shift in the equilibrium that favors the nonfunctional monomeric form. The unfolding is reversible, resulting in the formation of active dimers upon a temperature downshift. The pathogenic *S. typhimurium* uses this mechanism of direct temperature sensing to identify the mammalian host.

Bacterial Indirect Heat Sensors

As mentioned above, bacterial indirect heat sensors are molecular chaperones and proteases (DegS of *E. coli* described under Section 9.9, "The σ^E pathway") which modulate the activity of transcriptional regulators in response to the amount of non-native proteins within the cells. If their concentration is high, e.g., immediately after a heat shock, they allow expression of the heat shock genes under their regulation. The more non-native proteins have been removed by the HSPs, the more transcriptional activator molecules will return to their inactive (sigma factors) or active state (repressors) We will discuss five examples: the alternative sigma factors σ^{32} and σ^B and the repressor proteins HrcA, CtsR and HtpR.

In addition to its translation, the amount of active σ^{32} is also regulated at the level of stability. This regulation mechanism needs the DnaK chaperone team and the membrane-anchored ATP-dependent protease FtsH. While σ^{32} is subject to degradation at low temperature (30 °C), it becomes transiently stabilized after a heat shock to 42 °C. It is assumed that σ^{32} exists in two different conformations, an active and an inactive one, and there is an equilibrium within the cell between these two conformations. This equilibrium is influenced by the DnaK chaperone system, which recognizes the active conformation and allows binding of σ^{32}, favoring formation of the inactive conformation. Upon release from DnaK, the sigma factor is subject to degradation by FtsH. Immediately after a heat shock, the amount of non-native proteins increases dramatically in the cytoplasm, titrating the molecular chaperones including DnaK. This in turn leads to an increase in active σ^{32} which binds to the core RNA polymerase to allow transcription of the σ^{32}-dependent operons. The more non-native proteins have been either refolded through interaction with molecular chaperones or degraded by ATP-dependent proteases, the more DnaK will be available to convert σ^{32} from its active into its inactive conformation, which is subject to degradation by FtsH. This leads to a gradual shut-off of the enhanced expression of the genes of the σ^{32} regulon; and their transcription returns to a basal level which is two- to three-fold higher than the level observed in the absence of heat shock, provided the cells are further incubated at the high temperature.

The second example for an alternative heat shock sigma factor is σ^B of *B. subtilis*, whose activity is regulated in a rather complicated way. In the absence of heat stress, most molecules of σ^B are sequestered by the anti-sigma factor RsbW (for

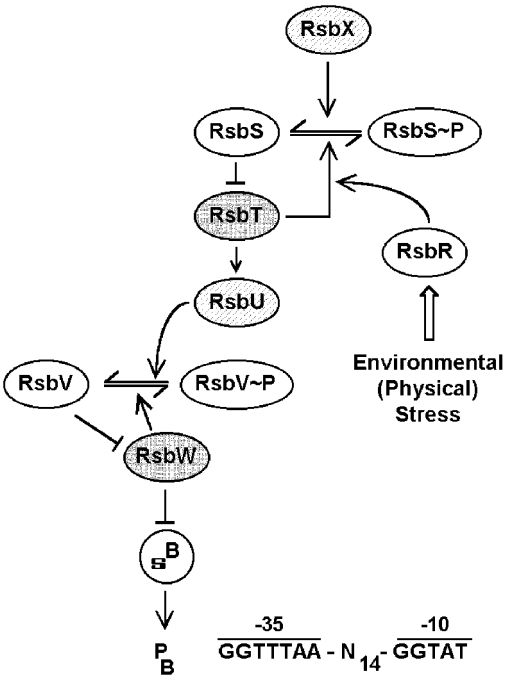


Fig. 9.3 The activity of the alternative sigma factor σ^B is modulated by different proteins in response to heat and other environmental stresses. See text for a detailed description. The consensus sequence recognized by σ^B is presented. W. Schumann **2003**, *Cell Stress Chaperones* 8, 207–217; Fig. 4, modified.

*r*egulation of *s*igma *B*; Fig. 9.3), resulting in a basal level of expression of the about 200 target genes. Besides binding σ^B, RsbW has kinase activity and phosphorylates the anti-anti-sigma factor RsbV in the absence of stress, keeping it in an inactive state. If *B. subtilis* cells experience an environmental stress factor such as a heat shock, the RsbU phosphatase becomes activated through RsbT. RsbT is kept inactive by interaction with RsbS. Upon a heat shock, RsbT acquires a kinase activity and phosphorylates RsbS, causing its dissociation from RsbT. The activation of RsbT is triggered by RsbR, while RsbX dephosphorylates RsbS later to turn off the heat shock response. It is tempting to speculate that both RsbR and RsbX either directly or indirectly sense non-native proteins within the cytoplasm.

One of the most widely distributed negative heat shock control mechanism is the HrcA/CIRCE system, which has been identified in more than 100 bacterial species, including Gram-positive bacteria (where the mechanism has been detected), proteobacteria and cyanobacteria. It has been discovered in *B. subtilis* and *C. acetobutylicum*, where sequence comparison of the *dnaK* and *groE* operons revealed the presence of a perfectly conserved inverted repeat, consisting of 9 bp separated by a 9-bp spacer downstream of the promoter. This inverted repeat was demonstrated to be involved in the heat shock response of the *dnaK* and *groE* operons both coding for molecular chaperones; and this finding prompted to designate it CIRCE (for *c*ontroling *i*nverted *r*epeat of *c*haperone *e*xpression; Fig. 9.4). The CIRCE element is recognized by the HrcA (for *h*eat *r*egulation at *C*IRCE) repressor which forms dimers in solution. The HrcA repressor of the hyperthermophile *T. maritima* has

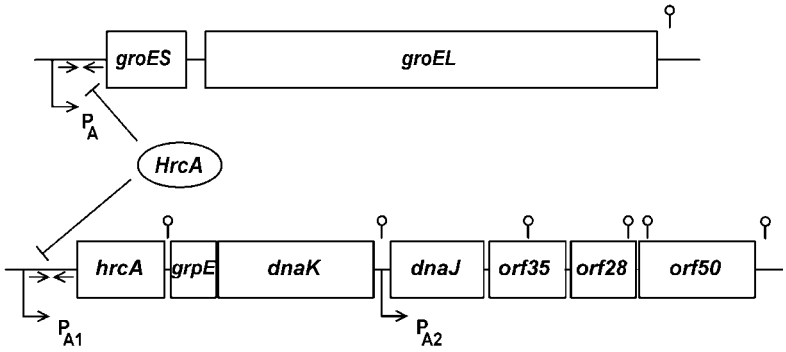


Fig. 9.4 The HrcA regulon of *B. subtilis*. The HrcA repressor controls expression of the bicistronic *groE* and the heptacistronic *dnaK* operon by binding to two inverted repeats symbolized by the inverted arrows. Promoters recognized by the houskeeping sigma factor σ^A and transcriptional terminators, symbolized by the haipin structure are indicated. W. Schumann **2003**, *Cell Stress Chaperones* 8, 207–217; Fig. 1, modified.

been crystallized and shown to form a dimeric structure. Each monomer is composed of three globular domains: an N-terminal winged helix-turn-helix domain, a central GAF-like domain and an inserted dimerizing domain.

How is the activity of the HrcA protein modulated by temperature? At low temperature (30–37 °C), HrcA binds to the two operators to allow a basal level of expression. After a sudden increase in temperature (48–50 °C), HrcA dissociates from its operators resulting in high level of transcription of both operons. What causes the repressor to dissociate from its operator? It can be assumed that a small effector molecule such as ppGpp is produced after a sudden temperature increase, interacting with the GAF domain and causing a conformational change, followed by dissociation of HrcA from its operator. About 10 min later, the repressor binds again to both operators to reduce transcription to close to the basal level, indicating release of the effector molecule and return of the repressor to its active conformation. Current data strongly suggest that the GroE chaperonin system plays an important role, but the molecular basis of this interaction has not yet been established. Since the crystal dimer structure may represent an inactive form of HrcA unable to bind to the CIRCE element, the GroE system has been suggested to interact with an about 25-amino-acid sequence near the C terminus, conserved in all the HrcA proteins, to convert the repressor into its active form. Why does it take about 10 min for the repressor to regain its active conformation? Immediately after a heat shock, the amount of non-native proteins increases the titrating of all available chaperone molecules including the GroE team. Two different mechanisms lead to an increase in the amount of free GroE able to take care of the inactive HrcA: First, expression of the *groESL* operon is increased after a heat shock, leading to an enhanced amount of both proteins; and second, the more non-native proteins that have been removed fom the cytoplasm, the more GroE chaperonins will become available. In summary, the non-native proteins titrate GroE after a heat shock, preventing reactivation of HrcA.

The second example of a transcriptional repressor is the CtsR (for *c*lass *t*hree *s*tress *r*epressor) protein first identified in *B. subtilis*, too. CtsR is a dimeric protein, which binds to a highly conserved heptanucleotide direct repeat, located upstream of the *clpC*, *clpE* and *clpP* operons (Fig. 9.5). At 37 °C, a basal level of active CtsR is present, while after a heat shock a rapid degradation of the repressor by the ATP-dependent ClpCP protease occurs. The *clpC* operon codes for the four genes *ctsR*, *mcsA*, *mcsB* and *clpC*, where McsA and McsB act as modulators of CtsR. McsB is a novel kind of tyrosine kinase using a guanidine kinase domain. At 37 °C, the kinase activity of MscB is inhibited by ClpC. After a temperature challenge, ClpC is titrated by denatured proteins, releasing MscB which becomes activated by MscA. MscB autophosphorylates and then phosphorylates MscA and CtsR, where all three proteins form a ternary complex. Formation of the ternary complex not only prevents the binding of CtsR to its operators, but also makes it prone to degradation by ClpCP. Here, MscB targets CtsR to ClpC. Furthermore, a cognate phosphatase has been identified (YwlE) that could antagonize the kinase activity of McsB.

The third example is the HspR (for *h*eat *s*hock *p*rotein *r*egulator) repressor identified in *S. coelicor*. This repressor has been shown to control expression of the tetracistronic *dnaK* operon (*dnaK*, *grpE*, *dnaJ* and *hspR*) and of *clpB* and *lon*. All three operons are preceded by an almost identical inverted repeat called HAIR (for *H*spR-*a*ssociated *i*nverted *r*epeat). The activity of HspR is modulated by the DnaK chaperone acting as a co-repressor, by binding to HspR at its operator site, and DnaK activates the HspR repressor. After a heat shock, DnaK is titrated by the non-native proteins, causing dissociation of HspR from its three operators. The more non-native proteins that have been removed, the more DnaK will be available to bind to HspR to down-regulate transcription of the three operons.

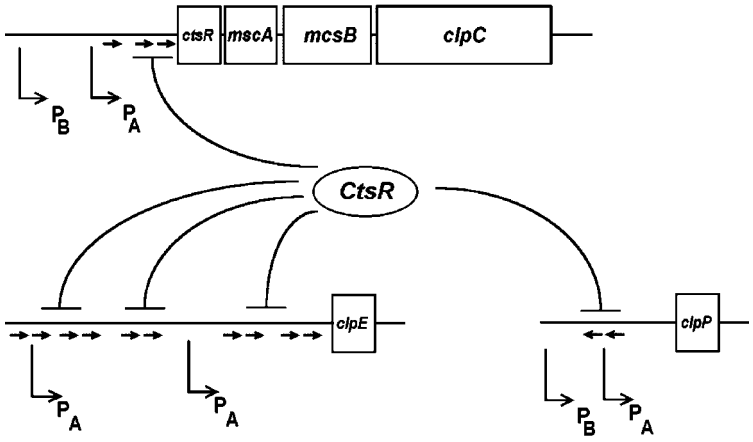


Fig. 9.5 The CtsR regulon. The CtsR repressor binds two direct repeats (symbolized by the arrows) present in multiple copies in front of the *clpC*, *clpE* and *clpP* operons. All three operons are under the control of the housekeeping sigma factor σ^A, the *clpC* and *clpP* operons in addition by the stress sigma factor σ^B. W. Schumann **2003**, *Cell Stress Chaperones* 8, 207–217; Fig. 5, modified.

9.3 The Cold Shock Response

A sudden downshift in temperature called cold shock drastically modifies all physico-chemical parameters of a living cell, influencing solute diffusions rates, enzyme kinetics, membrane fluidity and affecting conformation, flexibility and topology of DNA, RNA and proteins. Bacteria are exposed to a cold shock, for instance, when hail covers the earth during a thunderstorm after a hot sunny day or after excretion by mammals. Such a decrease in temperature results into two important consequences: (1) membrane fluidity decreases and (2) secondary structures in RNA and DNA are stabilized, resulting in reduced efficiencies of translation by reducing ribosomal movement on mRNA, transcription elongation and DNA replication. The first problem is overcome by increasing the degree of unsaturated fatty acids in the membrane phospholipids to attain more flexibility; and this response has been termed the *homeoviscous adaptation*. The deleterious effects mainly on translation are overcome by transient induction of the *cold shock proteins*.

Physiology After a Cold Shock

The cold shock response of exponentially growing cells can roughly be separated into three different phases. Phase I represents the initial but transient cold shock response (the *acclimation phase*, i.e., lag period of cell growth) that immediately follows cold exposure and, depending on the bacterial species, may take up to several hours, during which a profound reduction of the growth rate as well as reprogramming of protein synthesis is observed. When the mesophile *E. coli* is transferred from 37 °C to a temperature below 20 °C, cell growth stops for an acclimation period of one to several hours. During phase I, homeoviscous adaptation occurs and a dramatic reprogramming of gene expression takes place, leading to a transient increase in the rate of synthesis of a small set of *c*old *s*hock *p*roteins (CSPs) to overcome the deleterious effects of cold shock, whereas that of most of the other gene products is repressed. Many CSPs facilitate translation by adapting ribosomes to the lower temperature and preventing the formation of RNA secondary structures. During phase II, the recovery phase, cells start to grow significantly faster. These cells are considered to be cold-adapted and later enter phase III, the stationary phase.

The Homeoviscous Adaptation

As growth temperatures are lowered, the membrane undergoes a reversible change from a fluid (disordered) to a nonfluid (ordered) state. To compensate for the loss in fluidity, the level of unsaturated fatty acids is increased. This can occur either by the incorporation of fatty acids into phospolipids or by the introduction of double bonds into pre-existing saturated fatty acids by *desaturases*. Fatty acid desaturases are enzymes that introduce double bonds into the hydrocarbon chains

of fatty acids. *E. coli* adjusts its fatty acid composition in response to a lower growth temperature by increasing the amount of *cis*-vaccenic acid and decreasing the amount of palmitic acid incorporated into membrane phospholipids. At 37 °C, palmitic acid occupies position 1 of the phospholipid backbone, whereas palmitoleic acid is found only at position 2. As the growth temperature is lowered, *cis*-vaccenic acid competes with palmitic acid for position 1 of the newly synthesized phospholipid; and the increased rate of *cis*-vaccenic acid synthesis is evident within 30 s after cold shock. By what mechanism is the increase in *cis*-vaccenic acid regulated? β-Ketoacyl-ACP synthase II, encoded by the *fabF* gene, elongates palmitoleic acid to *cis*-vaccenic acid (C16:1Δ9 is converted to C18:1Δ11). This leads to an increase in unsaturated fatty acids and diunsaturated phospholipids which lower the melting point and allow a greater degree of flexibility compared to saturated phospholipids. While the synthesis of β-ketoacyl-ACP synthase II is not induced upon cold shock, the enzyme activity is enhanced at low temperature.

In contrast, *B. subtilis* contains a membrane-bound Δ5 acyl lipid desaturase (Δ5-Des), encoded by the *des* gene. A novel pathway, the Des pathway, responds to a decrease in growth temperature by enhancing the expression of the *des* gene encoding Δ5-Des. This Des pathway is uniquely controled by a two-component system composed of a membrane-associated kinase, DesK, and a soluble transcriptional activator, DesR. DesK is proposed to possess dual functionality, acting either as kinase or as phosphatase to phosphorylate or dephosphorylate DesR in a temperature-dependent manner. At 37 °C, DesK acts as a phosphatase to keep DesR in an inactive dephosphorylated form (Fig. 9.6). After a temperature downshift, the kinase-dominant state of DesK predominates, which uses Mg-ATP as phosphate donor for autophosphorylation at its invariant histidine residue (His188). Next, the phosphoryl group is transferred to Asp54 of the response regulator DesR, which binds to two target sequences preceding the *des* gene and allowing the RNA polymerase to initiate transcription of the *des* gene. Production of Δ5-Des results in the conversion of palmitate to *cis*-D5-hexadeceonate. This in turn is sensed by the DesK protein, activating its phosphatase activity. The phosphate is first transferred back to His188 and then relased into the cytoplasm as inorganic P_i. In summary, DesK acts as a thermometer to measure the temperature. DNA macroarray studies have shown that this thermosensor system regulates only the *des* gene and possibly its own components.

Cyanobacteria respond to a decrease in ambient growth temperature by desaturating the fatty acids of membrane lipids. All known cyanobacterial desaturases are intrinsic membrane proteins that act on acyl-lipid substrates. In the unicellular marine cyanobacterium *Synechococcus* sp. strain PCC 7002, three desaturases have been identified: desaturases Δ12 (*desA*), Δ15 (*desB*) and Δ9 (*desC*), which are responsible for the conversion of stearate to α-linolenate. In another unicellular cyanobacterium, *Synechocystis* sp. strain PCC 6803, the induction mechanism for *desA* has been elucidated. Here, two histidine kinases have been identified (Hik33 and Hik19) and one response regulator (Rer1). In contrast to *B. subtilis*, the membrane-bound Hik33 kinase controls expression of a number of cold-regulated genes. Reduction in membrane fluidity appears to affect the conformation of

Hik33, thereby activating its kinase function, which leads to an activation of the cytosolic histidine kinase Hik19. This second kinase transfers the phosphoryl group to the response regulator Rer1 mediating finally cold-induced expression of at least *desB*.

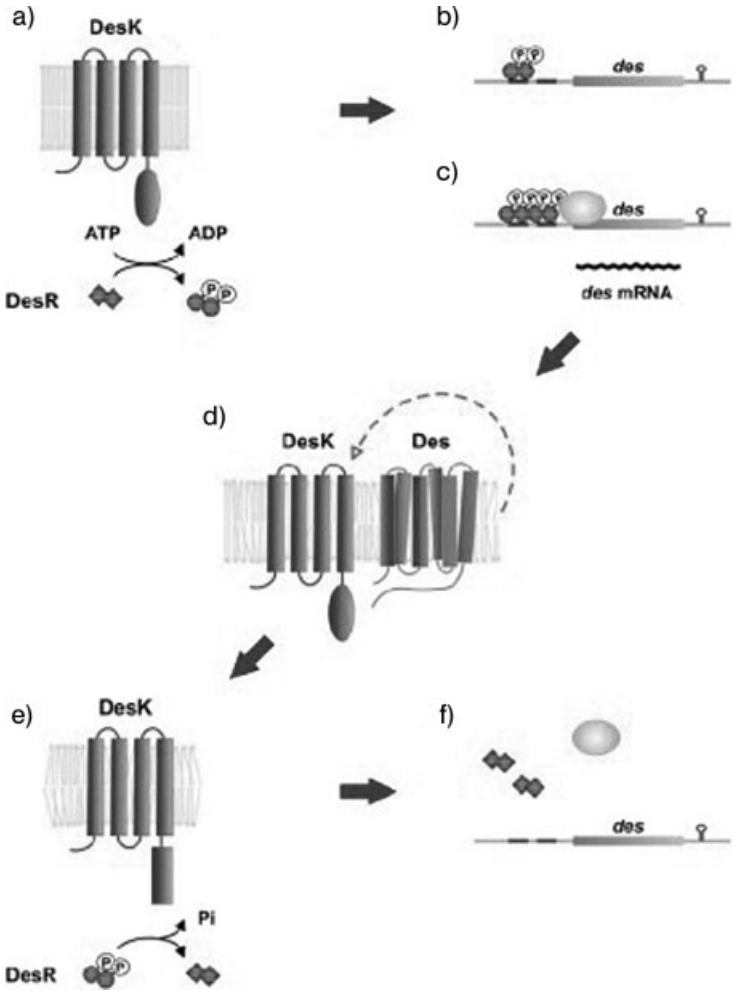


Fig. 9.6 Model of the Des pathway. (a) If cells experience a temperature decrease to 25 °C, this is sensed by DesK and leads to the activation of its kinase. DesK undergoes autophosphorylation and then transfers the phosphoryl group to the response regulator DesR. (b, c) The active DesR~P binds to DNA sites immediately upstream of the *des* promoter to initiate transcription by interaction with the RNA polymerase. (d) The desaturase inserts into the cytoplasmic membrane and introduces double bonds into acyl chains of the phospholipids which converts DesK from a kinase to a phosphatase. (e) Dephosphorylation of DesR~P results in turn-off of *des* transcription. M.C. Mansilla, et al. **2005**, *Arch. Microbiol.* 183, 229; Fig. 2. (This figure also appears with the color plates.)

Function of Cold Shock Proteins

Enterobacteria encounter sudden temperature downshifts as a result of excretion from animals. The cold shock genes which are induced are classified into types I and II. Class I genes are expressed at an extremely low level at 37 °C and are dramatically induced to very high levels after a shift to low temperatures ($\gg$10-fold), whereas class II genes are already transcribed at 37 °C and their induction factor is only a few-fold. Nine class I proteins have been identified so far: CspA (for *c*old *s*hock *p*rotein *A*), CspB, CspG, CspI, CsdA (*c*old *s*hock *D*EAD box protein *A*), RbfA, NusA, RNase R and PNPase. At 37 °C, CspA seems to exert a general role in chaperoning nascent RNA molecules, while CspE binds to poly(A) tails at the 3′ end of mRNAs and interferes with their degradation by both PNPase and RNase E. In the cold, the RNA chaperone activity of the Csp proteins may be used to control potentially unfavorable secondary structures of the RNA and to protect the cell against premature Rho-dependent terminations, an event expected to occur frequently if cold shock uncouples transcription and translation. CsdA (also called DeaD) belongs to the group of DEAD box RNA helicases. It bears a C-terminal extension (CTD) of ~20 kDa unique to its subfamily. It unwinds double-stranded RNA in the absence of RNA and assembles into degradosomes, with RNase E replacing the RhlB helicase under conditions of cold shock (cold shock degradosome). RbfA is a *r*ibosome *b*inding *f*actor, NusA is involved in termination and anti-termination of transcription, RNase R and PNPase are two 3′-specific exonucleases. Class II proteins include RecA, the initiation factors IF1, IF2 and IF3, the histone-like protein H-NS, the DnaK homolog Hsc66 and the 9 subunit GyrA of the DNA gyrase. Since not all cold shock genes seem to be induced by the same mechanism, these form a stimulon rather than a regulon.

Regulation of Cold Shock Genes

CspA is the major cold shock protein of *E. coli* and accounts for more than 10% of total protein synthesis during the acclimation phase. Its transcript contains a 159-nucleotide 5′ *un*translated *r*egion (UTR) which plays an important role in controling the stability of this RNA. At 37 °C, its half-life is extremely short ($t_{1/2} \approx 12$ s) and increases to about 70 min measured 30 min after a temperature downshift to 15 °C. RNase E is likely to be involved in the decay of the *cspA* transcript at 37 °C, since potential recognition elements for this RNase have been identified in the 5′ UTR (Fig. 9.7). A similar post-transcriptional regulation has also been postulated for the other *csp* genes (*cspB*, *cspG*, *cspI*), for *hns* and *infB*. PNPase seems to be responsible for the degradation of these transcripts during cold acclimation.

The translational apparatus of cold-shocked *E. coli* cells translates at 15 °C cold shock and, to a lesser extent, cold-tolerant mRNAs at a faster rate and to a higher level than non-cold shock mRNAs. How can this translational bias be explained? Two different elements seem to contribute to this bias: *cis*- and *trans*-acting elements. *Cis*-acting elements confer to the mRNA the capacity of being translated in the cold. This capacity is maximal in cold shock mRNAs (e.g., *cspA* and *hns*),

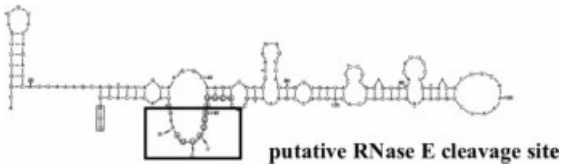


Fig. 9.7 Proposed secondary structure model for the 5′ untranslated *cspA* transcript. This secondary structure exposes a suggested recognition site for RNase E. The indicated base changes lead to a stabilization of the transcript. L. Fang, et al. **1997**, *Mol. Microbiol.* 23, 355–364; Fig. 7, modified.

somewhat less in the cold-tolerant *hupB* transcripts, and lowest in non-cold shock mRNAs (e.g., *hupA* and *cspD*; *hupA* and *hupB* code for the two subunits of the nucleoid-associated HU heterodimer). What do we know about the nature of the *cis*-elements? It seems that these are constituted by elements of secondary/tertiary structure rather than by sequence motifs as part of the primary structure. In the case of the *cspA* transcript, these elements are located in the upstream half (nucleotides 1–80) of the 5′ UTR. As to the nature of the *trans*-acting factors, the cold shock causes an imbalance of the initiation factor/ribosome ratio, and IF3 has been identified as being the most important *trans*-acting element. This initiation factor confers translational selectivity in the cold in favor of cold shock transcripts. Cold shock mRNAs exhibit an exceptionally high dependence of IF3 for their translation. As to the underlying mechanism, the first order isomerization of the ''30S preternary complex'' leading to ''30S initiation complex'' is kinetically controled by IF3; and this step may be rate-limiting, especially at low temperature. IF1 could potentiate the effect of IF3, while IF2 seems very unlikely to contribute as a *trans*-acting factor. An additional *trans*-acting factor is the nucleic acid binding protein CspA. CspA could stimulate mRNA translation by favoring unstructured transcript conformations, thereby facilitating their interaction with the ribosomes. This would be in line with its suggested role of an RNA chaperone. Additional factors may exist causing structural modifications of the ribosomes capable of enhancing the effects of the already known *cis*- and *trans*-acting elements.

C.O. Gualerzi, et al. **2003**, Transcriptional and post-transcriptional control of cold-shock genes, *J. Mol. Biol.* 331, 527–539.
M.H.W. Weber, M.A. Marahiel **2003**, Bacterial cold shock responses, *Sci. Prog.* 86, 9–75.

9.4 The Osmotic Shock Response

Bacteria are able to survive dramatic changes in extracellular osmolalities (osmolality is the osmotic pressure of a solution at a particular temperature, expressed as moles of solute per kilogram of solvent), as illustrated by three examples. Soil bacteria survive periods of low and high rainfall, uropathogens survive urine concentration and dilution and industrial bacteria tolerate concentrated nutrient solu-

tions as well as extracellular accumulation of metabolic products. Bacteria have evolved mechanisms allowing them to respond both actively and passively to changes in the osmolality in their environment. Alterations in the external osmolalities immediately trigger fluxes of water along the osmotic gradient, resulting either in swelling and bursting of the cell in *hypoosmotic* environments or in plasmolysis and dehydration under *hyperosmotic* conditions. The bacterial cell tries to avoid these devastating alternatives by using active processes to retain a suitable level of water and *turgor*. Turgor pressure is defined as a force pressing the cytoplasmic membrane towards the cell wall. Cells control their turgor by actively modulating solutes in their cytoplasm to allow the water content to be adjusted by osmosis. They amass ions and organic osmolytes in growth conditions of high osmolality (hyperosmotic shock); and they expel these compounds through mechanosensitive channels when external osmolality drops (hypoosmotic shock). In general, the total concentration of osmolytes within a cell is generally higher than in the surrounding environment, causing water to flow down its chemical potential into the cell. Turgor is essential for cell viability and is critical for growth, providing the mechanical force for expansion of the cell wall. Values of 3–5 bar (300–500 kPa) have been estimated for Gram-negative and approximately 20 bar (2000 kPa) for Gram-positive bacteria. The much higher value for the latter group is thought to reflect the large cytoplasmic solvent pool needed for growth of the multilayered cell wall. Water fluxes across the cytoplasmic membrane are accomplished by two different mechanisms: (1) simple diffusion and (2) water-selective channels. The *E. coli aq*ua*p*orin AqpZ mediates rapid and large water fluxes in both directions in response to sudden osmotic down- and upshifts. The hyperosmotic shock has been investigated in detail in *E. coli*, *B. subtilis* and in the photoautotrophic cyanobacterium *Synechocystis* sp. strain PCC 6803, the hypoosmotic shock in *E. coli*.

J.M. Wood **1999**, Osmosensing by bacteria: signals and membrane-based sensors, *Microbiol. Mol. Biol. Rev.* 63, 230–262.

J.M. Wood, et al. **2001**, Osmosensing and osmoregulatory compatible solute accumulation by bacteria, *Comp. Biochem. Physiol.* 130, 437–460.

9.4.1 The Hyperosmotic Shock Response

Many nonhalophilic bacteria respond to a sudden osmotic increase in their environment with a two-phase adaptation reaction. Initially, large amounts of K^+ are rapidly taken up by the cells through specific transport systems to compensate for the increase in the external salinity or osmolality. But the high intracellular concentration of K^+ exerts negative effects on protein function, DNA–protein interactions, and the synthesis of proteins. The second phase of osmoadaptation therefore frequently involves the synthesis or uptake of compatible solutes and the efflux of K^+. The adaptation mechanism towards hyperosmotic stress has been studied in several bacterial species; and we will deal here with those described for *E. coli* and *B. subtilis*.

The Adaptation Mechanism of *E. coli*

Adaptation to high-osmolarity stress is a two-step process. First, the cells take up K^+ and balance the increased K^+ concentration by accumulation of the counter-ion glutamate, the concentration of which is strongly increased by *de novo* synthesis. Other compounds, such as glutamine, γ-glutamylglutamine and glutathione, also contribute to the balance of the intracellular K^+ pool. This *primary response* to a high osmolarity environment results in the re-establishment of the osmotic pressure gradient across the inner membrane. Since high K^+ concentrations have deleterious effects on cell physiology and can inhibit the function of key enzymes, the *secondary response* results in replacement of much of the intracellular K^+ by so-called compatible solutes.

K^+ is the most abundant cation in the cytoplasm of *E. coli*. Its steady-state intracellular concentration increases from 0.15 M to 0.55 M when the osmolarity of the medium is increased from 100 mOsm to 1200 mOsm with solutes unable to permeate across the inner membrane. *E. coli* cells accumulate K^+ by three independent saturable systems: TrK, Kup and Kdp. Any one of these three systems is sufficient for growth in medium containing concentrations of K^+ in the low millimolar range. While the TrK and Kup systems are constitutive, the Kdp system is induced by high osmolarity.

The TrK transporter (K_m 0.3–3.0 mM) consists of the two homologous proteins TrkG (53 kDa) and TrkH (53 kDa), both integral inner membrane proteins with 12 transmembrane segments each. The TrkA (50 kDa) exists both in cytoplasmic and membrane-associated forms. Its association with TrkG and TrkH is required for K^+ uptake. The Trk systems are believed to be the earliest sensors of osmotic upshift in *E. coli*, though details how they register the stressful situation remain to be elucidated. The *kup* gene encodes a minor K^+ uptake system, and the 69-kDa Kup protein is composed of two domains. The first 440 amino acid residues appear to form an integral membrane protein that might traverse the inner membrane 12 times. The C-terminal 182 amino acid residues are predicted to form a hydrophilic domain located at the cytoplasmic side of the membrane.

The *kdpFABC* operon codes for an inducible, four-subunit, K^+-transporting P-type ATPase with a high affinity for K^+ and a low K_m (2 μM), found in *E. coli* and many other bacteria, and the two-component regulatory system KdpDE. The KdpA subunit is believed to span the cytoplasmic membrane ten times and to translocate K^+ in a process involving two sequentially occupied K^+ binding sites. The catalytic, membrane-associated KdpB subunit accepts phosphate from ATP during K^+ transport. The functions of the essential subunit KdpC and the hydrophobic peptide KdpF are unknown. The 98.5-kDa homodimeric KdpD protein is a membrane-bound sensor kinase consisting of a large N-terminal domain and a C-terminal transmitter domain interconnected by four transmembrane segments. KdpD contains three different activities: an autokinase, a KdpE-specific phosphotransferase and a KdpE~P-specific phosphatase. The 26-kDa KdpE response regulator is able to bind in both its unphosphorylated and its phosphorylated form to its cognate promoter region, but only KdpE~P is able to activate transcription. Ex-

pression occurs when growth becomes K^+ limited, suggesting that reduced turgor is the signal to express the *kdp* operon (the turgor control model).

The *secondary response* to high environmental osmolarity involves replacement of much of the intracellular K^+ by so-called compatible solutes. These osmotically active solutes can be accumulated to high intracellular concentrations without disturbing essential metabolic functions. They counterbalance high extracellular concentrations of osmolytes and maintain the turgor. A limited group of low-molecular-mass compounds are used as compatible solutes (Fig. 9.8): polyols (e.g., glycerol, trehalose), amino acids (e.g., proline), amino acid derivatives (e.g., taurine, β-alanine, ectoine), urea and methylamines (e.g., glycine betaine, proline betaine), where many of these compounds are found widely in nature.

E. coli cells synthesize the disaccharide trehalose in large amounts. Synthesis involves the condensation of glucose-6-phosphate and UDP-glucose to trehalose-6-phosphate via a trehalose-6-phosphate synthetase (OtsA). Trehalose-6-phosphate is then enzymatically converted to free trehalose by trehalose-6-phosphate phosphatase (OtsB). High osmolarity stimulates the transcription of *otsAB* and the enzymatic activity of the trehalose-6-phosphate synthetase is activated by K^+-glutamate. Glycine-betaine, an osmoprotectant of central importance for the adaptation of many bacteria to high-osmolarity environments, cannot be synthesized *de novo* by *E. coli*, but cells can scavenge its precursor molecules, choline and glycine betaine aldehyde, if present in the environment. Choline is taken up by BetT, a high-affinity (K_m = 8 μM) integral inner membrane choline transporter, which is energized by the proton motive force. Two enzymes are involved in the synthesis of glycine betaine from choline, a choline dehydrogenase (*betA*) and a glycine betaine aldehyde dehydrogenase (*betB*). BetA is a membrane-bound, oxygen-dependent enzyme that can catalyze the oxidation both of choline to glycine betaine aldehyde and of glycine betaine aldehyde to glycine betaine. The choline-sensing BetI repressor (21.8 kDA; HTH motif near the N-terminus) regulates *bet* gene expression negatively by binding to a 41-bp DNA sequence containing the –10 and –35 regions of both promoters. The *bet* gene cluster is induced under conditions of high osmolarity and is further stimulated by the presence of choline in the growth medium.

Another important osmoprotectant is proline. A large increase in the intracellular concentration of proline can be observed under high-osmolarity growth conditions, accomplished by uptake from the environment. *E. coli* codes for three dis-

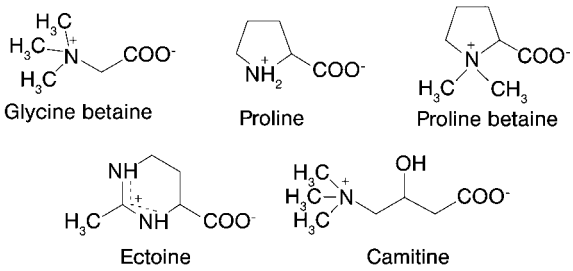


Fig. 9.8 Structures of important compatible solutes. J.M. Wood, et al. **2001**, *Comp. Biochem. Physiol.* 130, 437–460; Fig. 1.

tinct proline transport systems: PutP, ProP and ProU. The major proline permease, PutP (K_m = 2 μM), functions when proline is used as a nutrient, but plays no role in the uptake of proline at high osmolarity. In contrast, the activity of PutP is inhibited by osmotic stress. The remaining two transport systems, ProP and ProU function under conditions of osmotic stress and allow the intracellular accumulation of proline as an osmoprotectant. The ProP transporter (K_m = 0.3 mM) consists of a single polypeptide chain inserted into the inner membrane, is energized by proton motive force and is stimulated by high osmolarity, which might induce a conformational change in the ProP protein. ProP is assumed to contain 12 transmembrane segments where both the short N-terminal and the long C-terminal domains are exposed in the cytoplasm. The C-terminus contains a coiled-coil domain which is not essential for osmotic activation but tunes ProP to osmoregulate over a low osmolarity range. Cardiolipin, the concentration of which increases with extracellular osmolality, may contribute to ProP adaptation by altering bulk membrane properties or by acting as a ProP ligand. While ProP can sense and respond to osmotic upshifts, the soluble 26-kDa ProQ protein is required for full osmotic activation of ProP. ProQ binds to ProP homodimers created through C-terminal homodimeric coiled-coil formation causing a conformational change that stabilizes a fully active conformation of ProP. In addition to proline, ProP accepts taurine, ectoine, glycine betaine and structural analogs of glycine betaine as substrate. The third system, ProU, has a very high affinity for glycine betaine (K_m = 1.3 μM). The *proU* locus consists of three structural genes, *proV*, *proW* and *proX*, that are coordinately expressed under the control of one major osmoregulated promoter. The 37.6-kDa ProW hydrophobic protein is integrated into the inner membrane, while the 44.1-kDa ProV is associated with the membrane and is the energy-coupling component by hydrolyzing ATP. Both ProV and ProW are assumed to exist in the cell as homodimers. The third component, the 33.8-kDa ProX protein, is the glycine betaine binding protein and delivers this substrate to the ProW-ProV complex. At low osmolarity, the components of the ProU transport system are present in very small amounts, and uptake of glycine betaine is barely detectable. Expression of the *proU* locus is stimulated at the level of transcription upon a sudden osmotic upshock and can be triggered by a great variety of osmolytes that cannot permeate the cytoplasmic membrane; and the increased expression is maintained as long as the osmotic stimulus persists. The *proU* locus is preceded by a σ^{70}-type promoter with a poor match to the consensus sequence, including a suboptimal spacing of 16 bp between the –35 and –10 elements. Osmoregulation of *proU* depends on a DNA sequence extending approximately 200 bp 5′ to the –35 region which is associated with DNA curvature. It has been suggested that H-NS binds to this bent DNA.

Concomitant with the uptake of compatible solutes, K^+ is extruded from the cells. This involves two proteins, KefB (66 kDa) and KefC (68 kDa), two independent, glutathione-gated K^+ efflux systems present in *E. coli*. The two efflux systems are maintained in a closed state by glutathione; and the systems are fully activated by adducts formed by reaction of glutathione with electrophilic compounds, such as *N*-ethylmaleimide (NEM), methylglyoxal and chlorodinitrobenzene. Activation

of KefB and KefC provokes rapid K^+ efflux, accompanied by acidification of the cytoplasm and influx of Na^+ ions. Therefore, a major determinant of the sensitivity of *E. coli* cells to electrophiles is the cytoplasmic pH; and this can be modulated by the controled activation of KefA and KefC by glutathione adducts. KefB and KefC are proteins with 601 and 620 residues, respectively, and share 42% identical and 70% similar amino acids, suggesting gene duplication.

The Adaptation Mechanism of *B. subtilis*

The normal habitat of *B. subtilis* is the upper layers of the soil. Drought and rain drastically alter the osmotic conditions within this ecosystem. To survive and grow in its osmotically changing habitat, *B. subtilis* has developed highly integrated cellular stress adaptation reactions that are either part of σ^B-dependent general stress system or are specific to osmotic stress. For *B. subtilis*, turgor has been estimated at 19 atm (1900 kPa) and cells amasses ions and organic osmolytes when they are challenged by high osmolarity environments. A sudden upshock with 0.4 M NaCl triggers K^+ uptake, raising the K^+ pool from a basal level of approximately 350 mM to 650 mM. In a second step, the K^+ is replaced by compatible solutes, including proline, glycine betaine and choline, as already described for *E. coli*.

There is an anabolic pathway for the synthesis of proline under osmotic stress proceeding from glutamate and catalyzed by the ProA, ProJ and ProH enzymes (Fig. 9.9). ProJ and ProH are encoded by the *proHJ* operon, whose expression is enhanced upon growth in high-osmolality media. Furthermore, *B. subtilis* cells can achieve a considerable degree of osmoprotection by synthesizing glycine betaine, but the precursor choline must be taken up from the environment. *B. subtilis* possesses two high-affinity and osmotically regulated choline transporters (OpuB and OpuC; Fig. 9.9). While OpuB functions exclusively for choline transport, OpuC also recognizes a large number of preformed compatible solutes. Glycine betaine synthesis involves a two-step oxidation process catalyzed by the GbsA

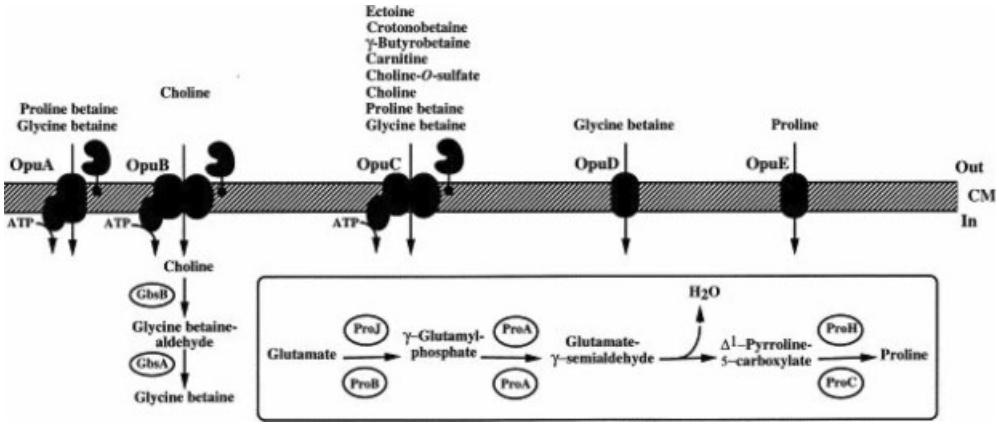


Fig. 9.9 Osmoprotectant transporters in *B. subtilis*. J.M. Wood, et al. **2001**, *Comp. Biochem. Physiol.* 130, 437–460; Fig. 2.

and GbsB enzymes. Both enzymes are encoded by the *gbsAB* genes forming a bicistronic operon negatively controled by the GbsR repressor protein. GbsR is a choline-sensing protein and mediates induction of the *gbsAB* and *opuB* operon. Besides proline and glycine betaine, *B. subtilis* can take up a variety of preformed compatible solutes, which are released into the ecosystem by root exudates and by decaying microbial, plant and animal cells via multiple transport systems with overlapping substrate specificity (Fig. 9.9). The proline transporter OpuE is member of the sodium/symporter family, the glycine betaine transporter OpuD of the BCCT (*b*etaine-*c*holine-*c*arnitine-*t*ransporters) family, while the OpuA and OpuC transport systems are members of the ABC family of transporters. In total, the Opu transporters can provide *B. subtilis* with at least eight preformed osmoprotectants and the precursor choline. While *opuB* expression is induced by choline, expression of *opuA*, *opuC*, *opuD* and *opuE* is not enhanced by the appropriate substrate but by increased medium osmolality. The mechanism(s) through which *B. subtilis* senses osmotic changes and transfers this information to the transcription apparatus is completely unknown. Both *opuD* and *opuE* genes are part of the σ^B regulon (see Section 9.2), but each has an additional, independently controled promoter responding to increases in medium osmolality.

9.4.2
The Hypoosmotic Shock Response

When *E. coli* cells are subject to hypoosmotic shock, this results in a substantial flow of water into the cytoplasm that can be equivalent to a 2- to 3-fold increase in the turgor (from 4 atm to approx. 11 atm; 400–1100 kPa) exerted from the cytoplasm on the membrane and the peptidoglycan wall, resulting eventually in cell bursting. To a certain extent, elevated turgor can be accommodated by the elasticity of the murein sacculus. When this capacity is exceeded, the cell must eliminate water-attracting osmolytes to avoid bursting. This is accomplished by so-called *mechanosensitive* (MS) *channels* acting as safety valves for the rapid release of these compounds when turgor rises beyond a threshold level. MS channels perform a remarkable transition between the closed and open state in response to membrane deformation arising from a sudden increase in the transmembrane pressure. Proteins that form these channels must have the capacity to maintain a tight seal in the closed state and not allow ion conduction. In the open state, the channels have been calculated at between 14 Å and 35 Å for *E. coli*. Such channels have been observed in all bacterial kingdoms, including Archaea. The major role for MS channels in cell physiology is protection of cell integrity when the cell is exposed to a dilute environment resulting in a hypoosmotic shock.

Three different types of MS channels have been identified, each with a characteristic unit conductance: McsL (high conductance), MscS (intermediate) and McsM (low). Each channel also exhibits a unique pressure threshold for activation, with the smallest channel opening at the lowest pressure. In *E. coli*, at least three sizes of channels with conductances ranging from 0.1 nS to 3.0 nS have been identified. These channels display little ion or solute preference. Transient

opening of the MS channels perturbs the normal gradients of the cell. While ATP, compatible solutes, glutamate and K^+ exit, Na^+ and H^+ enter. The changes in cytoplasmic Na^+ pool and pH support the notion that the solutes can move in both directions through an open channel. At neutral pH, *E. coli* cells recover from the transient opening of the channels. Two conditions generate high turgor in the cell: the accumulation of compatible solutes and the transfer of cells into media with low osmolarity. In both cases, the cells respond with the release of solutes from the cytoplasm to reduce the turgor pressure. Four different channels have been described in *E. coli*: three MS channels (MscL, McsK and MscS) and a water channel called aquaporin (AqpZ).

MscL, a 136-amino-acid protein, codes for the largest channel. MscL has the highest pressure threshold for activation of the channels and the largest conductivity (~3 nS). It seems to form either a pentamer or a hexamer in the inner membrane. Each monomer contains two transmembrane strands, a large periplasmic domain and small extensions at the cytoplasmically located amino and carboxyl termini. The size of this channel is such that the small protein thioredoxin (*trxA*; 11.7 kDa) can exit the cytoplasm via this channel during a hypoosmotic shock. MscS forms a small channel; and the MscS activity of bacterial protoplasts is eliminated by null mutations at two loci: *yggB* and *kefA*. While KefA is a large, multi-domain protein (1120 amino acids; 120 kDa), YggB consists of 286 amino acids highly similar at the sequence level to the last two domains of KefA. The two channels affected by the *kefA* and *yggB* null mutations exhibit similar conductance (1 nS) and similar pressure thresholds for activation, but differ in their abundance and in their properties. Whether *yggB* and *kefA* actually encode channels or are regulators of the expression of the MscS channel activities must await reconstitution studies. While *E. coli* mutants possessing either MscL or MscS channels do not express a phenotype, the double mutant lacking both MscL and YggB dies upon transfer from high to low osmolarity. Death takes place in the first few seconds after transfer and is accomplished by cell lysis. There is a minor channel, MscK, that does not seem to protect *E. coli* cells against severe osmotic transitions despite a relatively high conductance (300 pS).

While it was first assumed that the permeability of the lipid bilayer was sufficient to allow the rapid equilibration of water during rapid changes in the osmolarity of the medium, this view had to be changed after the discovery of homologs of the eukaryotic water channels termed aquaporins in *E. coli* and other bacteria. Expression of the *E. coli* aquaporin gene *aqpZ* is repressed as the osmolarity of the growth medium is increased; and *aqpZ* mutants do not exhibit a major phenotype. Rain, flooding and washout into freshwater expose *B. subtilis* cells to rapid and severe osmotic downshocks, causing a sudden entry of water into the cell followed by a drastic increase in turgor. As described for *E. coli*, mechanosensitive channels act as safety valves for eliminating water-attracting osmolytes to prevent bursting of the cells. The *mscL* gene is present in *B. subtilis*, and three YggB homologs that might function as MscS-type channels are also present.

M.D. Edwards, et al. **2004**, Gating the bacterial mechanosensitive channels: McsS a new paradigm? *Curr. Opin. Microbiol.* 7, 163–167.

9.5 The Oxidative Shock Response

Oxidative stress is defined as an intracellular surplus of prooxidants that originate in depletion of antioxidants or in the increase of *r*eactive *o*xygen *s*pecies (ROS). Primarily, ROS emerge as by-products from the normal respiratory activity of aerobically growing bacteria. Some of these by-products include singlet oxygen ($^1O_2$), peroxide radicals ($\bullet O_2$), hydrogen peroxide (H_2O_2) and hydroxyl radicals ($\bullet OH$). Although H_2O_2 is relatively stable, it can rapidly react with Fe^{2+} to produce highly reactive $\bullet OH$ radicals in a process described by the Fenton reaction. These $\bullet OH$ radicals can then react with DNA to produce a variety of DNA lesions. Reactions with guanine lead to 7,8-dihydro-8-oxoguanine (8-oxoG), which is the most common lesion. Exposure to ionizing radiation, redox-active compounds and metals may further enhance the intracellular amount of ROS, resulting in cell toxic levels if the concentration overexceeds the defense and repair capacities of the cell. Besides DNA, membrane lipids and proteins are susceptible to oxygen radical attacks. Oxidative modifications of proteins are assumed to impair or even inhibit a wide range of biochemical functions, such as interruption of metabolic pathways. Oxidation of proteins leads to modifications of amino acid sidechains, e.g., the generation of methionine sulfoxides, cysteine sulfonic acids, cysteine disulfide bonds or tyrosine crosslinks. Another evident alteration is the introduction of carbonyl groups in proteins at lysine, arginine, proline or threonine residues in a metal-catalyzed site-specific reaction.

E. coli cells utilize two transcription factors, OxyR and SoxR, to sense the oxidants and then induce various genes against oxidative stress that are involved in removing oxidants, repairing damaged cell components and maintaining reducing conditions in the cell. Whereas OxyR responds primarily to H_2O_2 and nitrosylating agents, SoxR is known to respond primarily to superoxide and nitric oxide.

9.5.1 The OxyR Regulon

The redox-sensing 34-kDa protein OxyR belongs to the LysR family of transcriptional regulators and responds to H_2O_2 and nitrosothiols. Since the level of OxyR protein does not change in cells treated with H_2O_2, this observation indicates that the protein is activated post-translationally. *In vitro* experiments have revealed that OxyR can exist in two forms: either the reduced or the oxidized form, where only the oxidized form is able to bind to DNA as a homotetramer (Fig. 9.10). The activated OxyR stimulates transcription by interacting with the αCTD of RNA polymerase. Oxidation of OxyR depends on the two cysteine residues, Cys199 and Cys208, which form a disulfide bond upon treatment of purified OxyR with H_2O_2. Nitrosothiols can release NO or transfer it to other molecules, and it was suggested that treatment of *E. coli* with nitrosothiols generates a distinct form of activated OxyR, which awaits experimental confirmation. OxyR also acts as a repres-

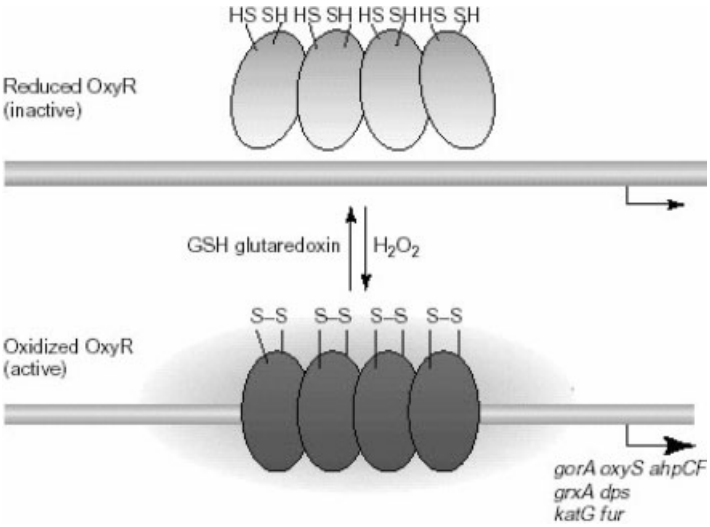


Fig. 9.10 The OxyR regulon. In the absence of H_2O_2, the OxyR transcriptional activator is present in its reduced form and unable to bind to DNA. The addition of reactive oxygen species leads to the oxidation of OxyR, followed by subsequent binding to promoter regions of its target genes to induce their transcription. P.J. Pomposiello, B. Demple **2001**, *Trends Biotechnol.* 19, 109–114; Fig. 3.

sor to autoregulate its own expression. Here, both the reduced and the oxidized forms are able to bind.

The OxyR-activated genes include *ahpCF* (alkyl hydroperoxide NADPH oxidoreductase), *fur* (an iron-binding repressor), *dps* (a DNA- and iron-binding protein), *gorA* (GSH reductase), *grxA* (glutaredoxin-1) and *katG* (peroxidase). Many of these proteins exert clear roles in antioxidant defense, such as removal of H_2O_2 by catalase from the cell and the protection of DNA by Dps from oxidative damage. The GSH-glutaredoxin-1 system contributes to restoring OxyR to the reduced state following oxidative stress. Additionally, OxyR activates the synthesis of a small ncRNA, *oxyS*. This RNA regulates 20 additional gene products indirectly by affecting either mRNA stability or translation efficiency.

9.5.2 The SoxRS Regulon

In *E. coli*, the SoxRS regulon is induced by the superoxide anion, a reactive oxygen species produced by reduction or by the action of redox-cycling compounds such as paraquat and menadione. In the presence of the appropriate oxidative stress signal, the 2Fe-2S centers of the constitutively expressed sensor, homodimeric SoxR, become oxidized, converting SoxR into an active transcription factor that initiates transcription of a single gene only, *soxS* (Fig. 9.11). The response regulator SoxS in turn leads to transcription activation of the regulon genes, the products of

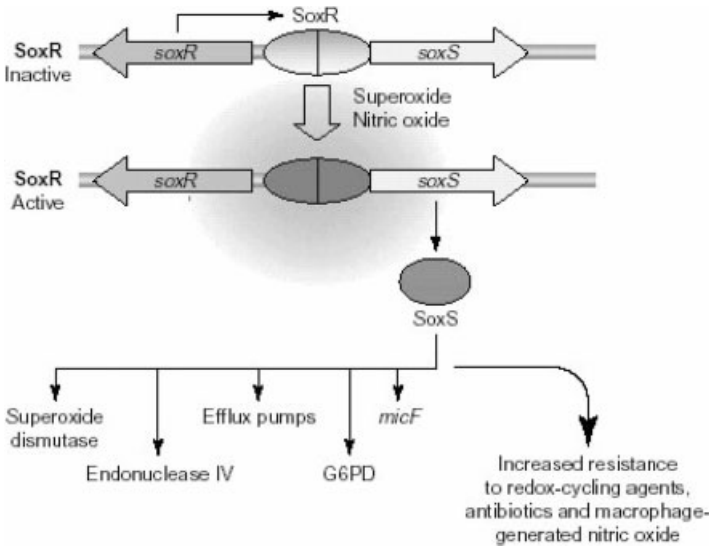


Fig. 9.11 Regulation of the SoxRS regulon. Upon exposure of cells to superoxide anions or nitric oxide, the transcriptional activator SoxR becomes oxidized and induces transcription of *soxS*. SoxS in turn activates transcription of a number of genes (a few are mentioned) enabling the cell to cope with the oxidative stress. P.J. Pomposiello, B. Demple **2001**, *Trends Biotechnol.* 19, 109–114; Fig. 1.

which provide the defense against the oxidative stress by removing the reactive oxygen species and repairing the damage caused by them.

SoxR is a 17-kDa transcriptional regulator of the MerR family. It forms a dimer in solution, with each monomer containing a [2Fe-2S] cluster which is not required for initial folding or for maintaining its structure or DNA-binding activity (Fig. 9.12). Instead, the [2Fe-2S] cluster undergoes reversible one-electron oxidation and reduction and thereby modulates its activity.When the [2Fe-2S] cluster of SoxR is in the fully oxidized state (Fe^{3+}-Fe^{3+}), SoxR can activate the transcription of its only known target gene *soxS*. When it is reduced by one electron (Fe^{2+}-Fe^{3+}), the ability to activate *soxS* transcription is lost. Thus, the Fe-S cluster serves as an elaborate redox-sensitive switch for SoxR activation to modulate *soxS* gene transcription. It has been estimated that the [2Fe-2S] clusters in SoxR are >90% reduced during aerobic growth. What is the underlying mechanism allowing SoxR to act as an activator? Even in the absence of stress, SoxR binds the *soxS* promoter to a site between the –10 and –35 elements of the *soxS* promoter, a rather unusual site for a transcriptional activator. Another unusual fact is that the spacer region between the –10 and –35 elements is 19 bp, compared with 17±1 bp spacer typical for *E. coli* housekeeping promoters. Furthermore, the RNA polymerase can bind to the SoxR-promoter complex, but can form only the closed complex. Upon exposure to superoxid anions, oxidation of the Fe-S cluster results in remodeling of the promoter compensating for the extended spacer region which stimulates formation of the open complex.

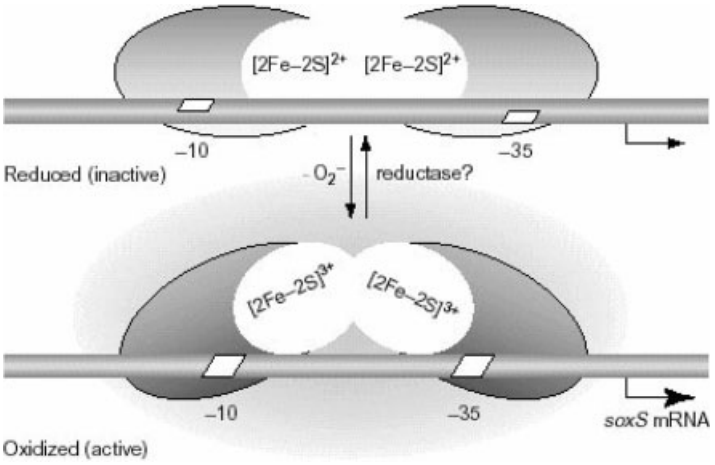


Fig. 9.12 Activation of SoxR needs oxidation of the two $[2Fe\text{-}2S]^{2+}$ clusters. In the absence of superoxide anions, the SoxR protein binds to the DNA in its reduced form and is unable to stimulate transcription of the *soxS* target gene. Oxidation of both $[2Fe\text{-}2S]^{2+}$ clusters causes a conformational change now enabling SoxR to act as a transcriptional activator. P.J. Pomposiello, B. Demple **2001**, *Trends Biotechnol.* 19, 109–114; Fig. 2.

The redox potential of SoxR at pH 7.6 has been estimated to be about –285 mV. More than 40% and up to 95% of SoxR exists as a form containing reduced [2Fe-2S] during aerobic growth. Since the cytoplasmic redox potential of *E. coli* has been estimated to be in the range of –260 mV to –280 mV, it is very likely that SoxR *in vivo* is actively maintained in the reduced state. How is SoxR maintained in its reduced state? Genes constituting the SoxR-reducing system were mapped in the *rsxABCDGE* (*r*educer of *Sox*R) operon and in the *rseC* gene, but details of the reduction mechanism are not yet known. As to SoxS, this protein is unstable since it is a target of the Lon protease, which needs access to the N-terminus of the SoxS polypeptide.

P.J. Pomposiello, B. Demple **2001**, Redox-operated genetic switches: the SoxR and OxyR transcription factors, *Trends Biotechnol.* 19, 109–114.
J. Green, M.S. Paget **2004**, Bacterial redox sensors, *Nat. Rev. Microbiol.* 2, 954–966.

9.6 pH-regulated Genes

As already mentioned, bacteria display an amazing capacity to survive and grow in extremely hostile environments. As an example, *E. coli* and *S. typhimurium* manage to grow in minimal medium ranging between pH 5.0 and pH 8.5, which represents a more than 3000-fold range in H^+ ion concentration. Both bacterial species can also survive from pH 4 to pH 9 (a 100000-fold range) for extended

periods of time. However, if the cells are first allowed to adapt to moderate acid or alkaline conditions before testing their limits, they can even survive over a 1000000-fold range of H^+ ion concentrations. This capability to be able to adapt can be very important in both natural and pathogenic situations, where pH conditions can fluctuate dramatically, and has been designated *acid* or *alkaline* tolerance. To give an example, enteric bacteria encounter a wide range of external pHs in their natural habitat, the human digestive tract. Colonization of the intestine requires transient survival through the stomach at pH 1–2 (fasting) or pH 2–7 (transiently, during feeding), as well as exposure to pancreatic secretions at pH 10, followed by growth and persistence in external pHs in the range pH 5–8. Growth at a pH substantially higher or lower than the cytoplasmic pH 7.6 induces protective responses with two fundamental aims: First, to maintain *internal pH homeostasis* and second, to prepare the cell to survive future exposure to more extreme pH conditions (below pH 5 or above pH 9; acid and alkaline tolerance) that no longer permit growth. The effects of pH are complex, because they intersect with other environmental factors such as oxygenation, growth phase and various metabolites.

How do cells manage to maintain an internal pH different from the external pH, called internal pH homeostasis? *E. coli* and *S. typhimurium* try to maintain a stable internal pH of pH 7.6–7.8 over a wide range of external pH conditions. It seems to be based on the low proton conductance of biological membranes and the presence of proton-driven transporters which either transport protons into the cell at basic external pH or extrude them out at acidic external pH. Four major systems located in the cytoplasmic membrane are responsible for controling proton flow into and out of the cell:

- H^+-coupled ion transport systems, which exchange ions such as K^+, Na^+ and H^+;
- H^+-F_1F_0 ATPases, which couple movement with the synthesis and hydrolysis of ATP;
- electron transport chains, which are responsible for H^+ efflux;
- genes encoding decarboxylases and deaminases.

Na^+/H^+ Antiporters

Proper intracellular concentrations of Na^+ (low) and K^+ (high) are crucial for the functioning of many intracellular reactions. Consequently, proton-driven antiporters are important for controling the intracellular concentration of these two ions. Na^+/H^+ antiporters help to maintain the internal pH, protect cells from excess sodium and have been implicated in Na^+ tolerance as well as acidifying the cytoplasm when cells grow in alkaline conditions. In *E. coli*, genes for the three distinct antiporters *nhaA*, *nhaB* and *chaA* have been characterized. The NhaA and NhaB antiporters of *E. coli* specifically exchange Na^+ or Li^+ for H^+. NhaA is required for adaptation to high salinity, resistance to Li^+ toxicity and growth at alkaline pH in the presence of Na^+. The 47-kDa NhaB protein contains 12 putative

transmembrane helices and confers a limited sodium tolerance to bacteria, but becomes essential in the absence of NhaA. The *chaA* gene codes for Ca^+/H^+. ChaA was reported to be a calcium/proton antiporter, but its physiological role was to transport sodium ions. Of the Na^+/H^+ antiporters, ChaA is unique in that it also shows pH-independent Ca^+/H^+ antiporter activity. ChaA is also regulated by Mg^{2+}, which inhibits both its Na^+/H^+ and Ca^+/H^+ antiporter activity.

H^+-F_1F_0 ATPases

H^+-F_1F_0 proton-translocating ATPase activity couples H^+ movement into the cell with ATP synthesis, or extrudes protons at the expense of ATP. The role of these H^+ ATPases in conferring acid tolerance to lactococcal, enterococcal and streptococcal species is well documented. When the intracellular pH drops, protons are extruded by the F_1F_0 ATPase accompanied by the electrogenic uptake of K^+. In *E. coli*, this enzyme is generally thought to be constitutive and unregulated by environmental factors. *E. faecalis* possesses an F_1F_0 ATPase that does not function in ATP synthesis, but does extrude protons in response to acidification. As the internal pH of the cell decreases, the synthesis and activity of the F_1F_0 ATPase increases, leading to the extrusion of protons, which restores the internal pH to neutrality.

Electron Transport

Electron transport chains constitute the third major system responsible for proton flow, where qualitative and quantitative changes occur with changing growth conditions. Important parameters are the type and availability of carbon sources, growth stage, the nature of electron acceptors and O_2 tension. The composition of the *E. coli* electron transport chain varies as a function of the O_2 level. Under conditions of low O_2, the low-affinity cytochrome *o* oxidase complex (*cyoABCDE*) is repressed and the high-affinity cytochrome *d* system (*cydAB*) is induced. Repression of the *cyo* operon is mediated by Fnr and ArcA, two important regulators responsible for the activation and repression of several genes as a function of O_2. This operon is also subject to pH regulation. Under aerobic conditions, expression varies four-fold between pH 6.0 and pH 7.0, while under anaerobic conditions, expression is completely repressed between pH 5.5 and pH 7.5 in an *fnr*$^+$ strain; but in the absence of Fnr, maximal repression is only observed below pH 6.0.

Genes Encoding Decarboxylases and Deaminases

In *E. coli*, the biosynthetic arginine and ornithine decarboxylases are constitutively produced in low amounts and unaffected by pH fluctuations. In contrast, the biodegradative decarboxylase genes including those for lysine (*cadA*), ornithine (*speF*) and arginine (*adi*) are induced by acid pH. These enzymes help the bacteria to survive pH stress by neutralizing their environment pH or by maintaining the internal pH. The 78-kDa lysine decarboxylase converts exogenously supplied lysine to cadaverine, an alkaline product that, upon secretion from the cell, neutralizes an

acidic environment. An antiporter for lysine/cadaverine is encoded by *cadB*. Expression of the *cadBA* operon is dependent on the CadC positive activator, with a putative membrane-spanning domain and a putative DNA-binding domain. CadC is believed to act as an environmental sensor and as a transcriptional regulator. According to the actual model, an external sensory domain of the molecule is exposed to the periplasmic space, where the external proton concentration can be detected directly or indirectly, while the N-terminal DNA-binding domain acts in turn on transcription at the *cadAB* promoter.

At pH 9, tryptophan deaminase (TnaA) is induced to a high level, becoming one of the most abundant proteins observed. TnaA may reverse alkalinization by metabolizing amino acids to produce acidic products. Why is this particular deaminase elevated to one of the most abundant proteins in the cell? TnaA deaminates not only tryptophan, but also serine and cysteine, producing pyruvic acid, which can be further degraded anaerobically to acetic and formic acids. Thus, induction of TnaA could offer a particularly effective means of neutralizing excess alkali. The stress conditions were designed to force *E. coli* to grow at the extreme limits of tolerance for pH.

H.K. Hall, et al. **1996**, Molecular responses of microbes to environmental pH stress, *Adv. Microbiol. Physiol.* 36, 229–272.

9.6.1
Acid Stress

Acid-resistance Systems

In *E. coli*, four acid-resistance (AR) systems have been described allowing cells to survive dilution into minimal media without glucose (AR1), with glucose and glutamate (AR2) or with glucose and arginine (AR3). The fourth, much less efficient system (AR4) relies on lysine. The first system (AR1) becomes apparent when cells are grown to stationary phase in LB medium that has been buffered to pH 5.5. These acid-adapted cells survive dilution into minimal media at pH 2.5, which rapidly kills nonadapted cells. The stationary phase specific sigma factor σ^S and CRP are required to develop acid tolerance, indicating that the AR1 system is glucose-repressed. Which σ^S- and CRP-dependent gene products are specifically needed remains unknown. Furthermore, the F_1F_0 proton-translocating ATPase is involved, but it is not clear whether this protein complex produces ATP to fuel AR1 or to directly extrude protons from the cytoplasm. *E. coli* cells grown in LB with glucose cannot survive exposure to pH 2.5 in minimal media unless glutamate is supplied.

The acid-resistance systems AR2 and AR3 depend on a decarboxylase/antiporter pair. The glutamate decarboxylase isoenzymes GadA and GadB and the arginine decarboxylase AdiA are pyridoxal phosphate-containing enzymes that replace the α-carbonyl groups of their amino acid substrate with a proton recruited from the cytoplasm. The decarboxylation reaction leads to the production of γ-amino butyric acid (GABA) in the case of GadA and GadB and agmatine in the case of the

AdiA enzyme, and CO_2 in all three cases (Fig. 9.13). The cognate antiporters are GadC for glutamate and AdiC for arginine; and they expel the decarboxylation products in exchange for importing new amino acid substrates. These mechanisms suggest that removal of intracellular protons causes the internal pH to increase. This has been proven by measurements of the internal pH. At pH 2.5, cells relying on the glutamate system maintain an internal pH of about pH 4.2, whereas cells dependent on arginine maintain an internal pH of approximately pH 4.7. How widespread are decarboxlase-dependent acid resistance systems? Glutamate-dependent systems have been described for *Lactococcus*, *Listeria* and *Shigella* and are assumed to be present in other bacterial species, based on *in silico* analysis of their genomes.

What is known about the genetic regulation of these systems? To date, 11 regulatory proteins are known to affect induction of AR2. The major activator, GadE, binds to a 20-bp sequence termed the *gad* box situated 63 bp upstream of the transcriptional start sites of *gadA* and *gadBC*. The *gadE* gene itself is subject to multiple activation circuits. One of these circuits includes the membrane-bound sensor kinase EvgS, the response regulator EvgA and the second transcriptional regulator YdeO. Furthermore, GadE autoactivates its own transcription. All three regulators EvgA, YdeO and GadE bind upstream of the *gadE* gene to stimulate its transcription. Both EvgA and YdeO are important for induction during the exponential

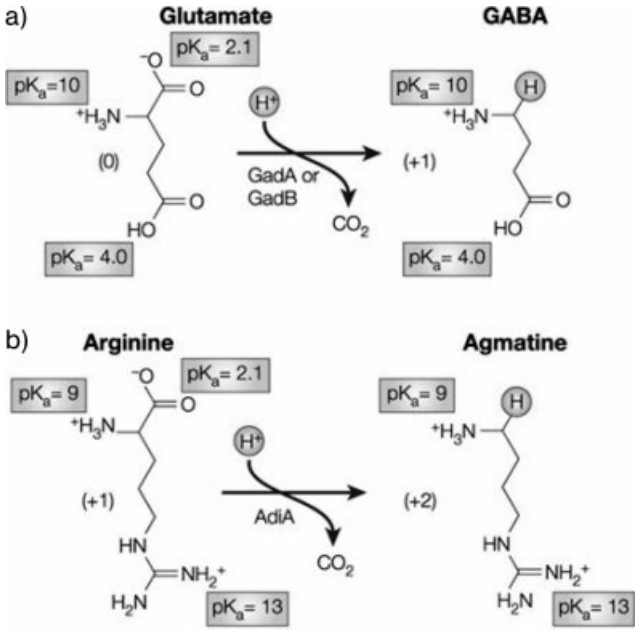


Fig. 9.13 Consumption of protons during decarboxylation of: (a) glutamate to GABA and (b) arginine to agmatine. pKa values for ionizable groups are indicated, numbers in parenthesis indicate the charge of the compound. J.W. Foster **2004**, *Nat. Rev. Microbiol.* 2, 898–907; Fig. 1.

growth phase in minimal medium. During the stationary phase, a second circuit leads to the activation of *gadE*. This pathway includes CRP, σ^S and the two activator proteins, GadX and GadW. The genes *gadX* and *gadW*, located downstream of *gadA*, are transcribed by independent promoters, and their proteins exert a dual regulatory function. Under some conditions, they activate expression of *gadA* and *gadBC*; and under different conditions, they bind to the *gad* boxes to inhibit transcription. Regulation of expression of the arginine decarboxylase/antiporter is less complex. The genes *adiA* and *adiC* are separated by *adiY* coding for a transcriptional activator of *adiA* and *adiC*, but the environmental conditions leading to its expression have not yet been identified. A second proven activator of these genes is CysB, which might also act as a sensor of anaerobic conditions and low pH values in complex media where *adiAC* are transcribed.

Sensors of Extracellular Acidity in *Agrobacterium tumefaciens*

Acid conditions play an important role in *Agrobacterium*–plant interactions. Expression of the *vir* genes is controled primarily through the VirA/VirG two-component system and depends on external acidification following release of acids at the wound of the plant. Two distinct pH-mediated responses of *vir* gene induction have been identified: (1) the acid-dependent transcription of *virG* and (2) the VirA/VirG-dependent transcription of the *vir* operons; and both responses are maximally induced at acidic pH. The ChvG/ChvI two-component signal transduction system is involved in the regulation of acid-inducible genes, residing on the circular and linear chromosomes, respectively, and the Ti plasmid. It is assumed that ChvG acts as a global sensor protein that directly or indirectly senses the acidity of the environment. A comparison of the *Agrobacterium* ChvG sequence with genomes of other α-proteobacteria revealed that the entire ChvG length is homologous to the corresponding protein sensors in *Sinorhizobium*, *Mesorhizobium*, *Brucella*, *Bartonella* and *Caulobacter*, underlining the significance of this two-component system. But it has to be stressed that there is no experimental proof that the sensor kinases in these other bacterial genera are actually sensing acid stress.

J.W. Foster **2004**, *Escherichia coli* acid resistance: tales of an amateur acidophile, *Nat. Rev. Microbiol.* 2, 898–907.

9.6.2
Alkaline Stress

All prokaryotes maintain a sodium concentration gradient directed inward and a constant intracellular pH around neutral. Na^+/H^+ antiporters play a primary role in these homeostatic mechanisms and are ubiquitous proteins inserted in the cytoplasmic membranes of cells and in the membranes of many organelles. Within *E. coli* and other enteric bacteria, antiporters encompass the primary systems responsible for adaptation to growth in conditions of high Na^+ concentrations and

varying pH. In *E. coli*, genes for the three distinct antiporters *nhaA*, *nhaB* and *chaA* have been characterized. The NhaA and NhaB antiporters of *E. coli* specifically exchange Na^+ or Li^+ for H^+. NhaA is required for adaptation to high salinity, resistance to Li^+ toxicity and growth at alkaline pH in the presence of Na^+. The 47-kDa NhaB protein contains 12 putative transmembrane helices and confers a limited sodium tolerance to bacteria but becomes essential when the lack of NhaA limits growth. Of the Na^+/H^+ antiporters, ChaA is unique in that it also shows pH-independent Ca^+/H^+ antiporter activity. ChaA is also regulated by Mg^{2+}, which inhibits both its Na^+/H^+ and Ca^+/H^+ antiporter activity. The *cha* operon consists of three genes, *chaA*, *chaB* and *chaC*, where both ChaB and ChaC are proposed to be regulators of *chaA*. However, the biological function for either remains to be established.

The best studied alkali-inducible system in *E. coli* is that of the sodium proton antiporter NhaA. While expression of the *nhaA* gene is low in the acid range, it increases 7-fold over the external pH range pH 7.0–9.0. The membrane-bound NhaA protein detects external pH directly; and at pH 8.4 it pumps two H^+ in for every Na^+ pumped out. This antiporter is regulated by two different mechanisms: at the level of activity and by a transcriptional activator. The NhaA activity is stimulated about 2000-fold by high pH and has been assumed to sense Na^+ and/or H^+, which influence the activity. NhaA has an optimum activity at alkaline pH. Transcription of *nhaA* is stimulated by the 34.2-kDa activator NhaR belonging to the LysR family, an integral membrane protein with two transmembrane segments. The *nhaA* gene has a dual mode of regulation of transcription, each involving a different promoter. During the logarithmic phase of growth, the expression of *nhaA* is positively regulated by NhaR, a member of the LysR family. Na^+ is the inducer and P1 is the Na^+-specific promoter which is transcribed by σ^{70}. In the stationary phase, σ^S transcribes *nhaA* via P2 in a fashion which is independent of Na^+ and NhaR.

Alkali-inducible Genes in *Bacillus subtilis*

In *B. subtilis*, about 80 genes are induced at least 4-fold by an alkali shock. While about 60 of these genes belong to the σ^W regulon, regulation of the other genes remains elusive, but two of them seem to be preceded by a riboswitch. In the absence of alkaline stress, the ECF sigma factor σ^W is sequestered by the anti-sigma factor RsiW (for *r*egulation of *si*gma *W*). RsiW consists of three different domains: the N-terminal domain is exposed in the cytoplasm and binds σ^W, the central domain anchors the protein in the cytoplasmic membrane and the C-terminal domain is exposed on the outside and is assumed to sense the alkali stress directly or indirectly (Fig. 9.14). Release of σ^W after a sudden increase of the external pH from pH 7.0 to pH 8.9 involves three different proteases acting successively. A so far unknown protease carries out a proteolytic cut in the external domain (called site-1 proteolysis), followed by a second cut within the membrane-anchor through the RasP protease (site-2 proteolysis). This causes release of the N-terminal polypeptide chain into the cytoplasm with the sigma factor still bound. Next, the new

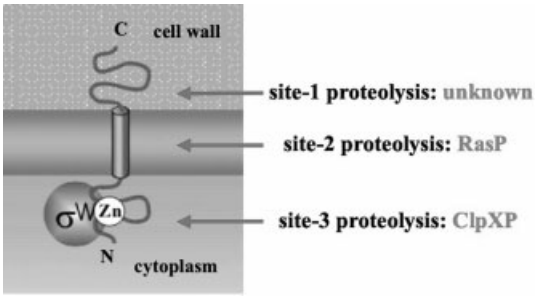


Fig. 9.14 Release of σ^W after alkali stress depends on three different proteases. The anti-sigma factor RsiW is anchored in the cytoplasmic membrane, while its C-terminal domain is exposed on the outside, acting directly or indirectly as stress sensor, and its N-terminus is located within the cytoplasm and sequesters the sigma factor σ^W. After alkali challenge, a still unknown protease carries out site-1 proteolysis within the C-terminal domain, followed by site-2 proteolysis within the membrane anchor. The remaining polypeptide chain is released into the cytoplasm and is recognized by the ClpXP protease, degrading it to small peptides and causing the release of σ^W. The free σ^W binds to the RNA polymerase core enzyme to transcribe the ca. 60 genes of the σ^W regulon.

C-terminus of the shortened anti-sigma factor is recognized by the ATP-dependent ClpXP protease and degraded into peptides. The site-3 proteolysis step finally causes release of σ^W into the cytoplasm where it binds the core enzyme of the RNA polymerase to transcribe the genes of the sigma-W regulon. It is an open question whether either RsiW, the first protease or another so far unknown protein acts of an alkali sensor.

9.7 Metabolic Stress: the Stringent Response

If cells are starved for an amino acid, the ribosomes will stall whenever they encounter a codon for the missing amino acid. Under these conditions, rRNA synthesis can continue, since RNA does not contain amino acids. However, in *E. coli* and probably other bacterial species, the synthesis of rRNA and tRNA ceases when an amino acid is missing. This coupling of the synthesis of rRNA and tRNA to translation is termed the *stringent response* and saves energy. There is no reason to make new ribosomes and tRNA when at least one amino acid is absent from the cells.

Detection of the Stringent Response

The stringent response adjusts anabolism with available resources. It serves as a control mechanism that reduces the cellular protein synthesis capacity when substrates for protein synthesis get scarce and maintaining a high level of protein synthesis machinery would be a waste of energy. The stringent response was detected as an inhibition of stable RNA (tRNA, rRNA) synthesis when *E. coli* cells

were starved for amino acids. Then, a mutant was isolated which did not exhibit the stringent response and continued to synthesize stable RNA during amino acid starvation. The gene inheriting the mutation was called *relA* (for *rel*axed *A*). Later, it was found that not only amino acid starvation, but a multitude of nutritional limitations, such as nitrogen, phosphorus and carbon limitations and a variety of other stresses (including hyperosmotic and oxidative stress, 10% ethanol and $CdCl_2$) will lead to induction of the stringent response. All these events lead to an increase in an effector molecule, ppGpp. Two enzymes have been identified being able to synthesize ppGpp: RelA and SpoT.

Synthesis of the Alarmone ppGpp and its Mode of Action

As mentioned, the stringent control is triggered by an effector molecule, the *alarmone* ppGpp (guanosine tetraphosphate, also known as magic spot). First, the RelA (pppGpp synthetase I) or SpoT (pppGpp synthetase II) enzymes catalyze production of pppGpp using GTP (or GDP) and ATP (Fig. 9.15). Next, they convert pppGpp into ppGpp, the active alarmone. RelA is associated with ribosomes and produces ppGpp in response to uncharged tRNA in the ribosomal A-site. SpoT, in contrast, is responsible for the accumulation of ppGpp in response to the stresses and nutrient limitations mentioned above, apart from amino acid starvation. In addition, SpoT is able to hydrolyze ppGpp. Two additional enzymes are involved in regulation of the stringent response, where Gpp hydrolyzes pppGpp to ppGpp and Ndk restores the pool of GTP by phosphorylating GDP to GTP. All cellular RelA is associated with the 50S ribosomal subunits, but due to the low abundance of RelA only 0.5–1.0% of the ribosomes have bound RelA. RelA interacts with a proline-rich helix in the N-terminal part of the ribosomal protein L11 encoded by the *rplK* (or *relC*) gene (Fig. 9.15); and this interaction is necessary for activation of the RelA activity.

ppGpp binds to the RNA polymerase and affects transcription. Mutations affecting ppGpp binding to the polymerase have been mapped in the β, β′ and σ^{70} subunits, while crosslinking studies revealed interaction with β and β′ subunits. ppGpp can act both as a negative and as a positive regulator of transcription. σ^{70}-dependent genes involved in cell growth and proliferation are negatively regulated and genes implicated in stress defense and maintenance are positively controled by ppGpp. In both cases, the alarmone seems to cause destabilization of the open complexes, but with different outcomes. While in the first case, destabilization causes dissociation of the RNA polymerase (especially at rRNA promoters), in the second case, it causes promoter clearance. The negatively regulated promoters are distinguished by several characteristic features: (a) the presence of a so-called GC-rich *discriminator* sequence between the –10 promoter region and the start site of transcription, (b) a shortened 16 bp spacer between the –35 and –10 boxes of the promoter, (c) a non-canonical sequence of the –35 region and (d) the formation of short-lived open complexes.

Recently, a protein has been described, DksA, assumed to contribute to both the negative and positive effects of ppGpp on transcription. DksA binds RNA

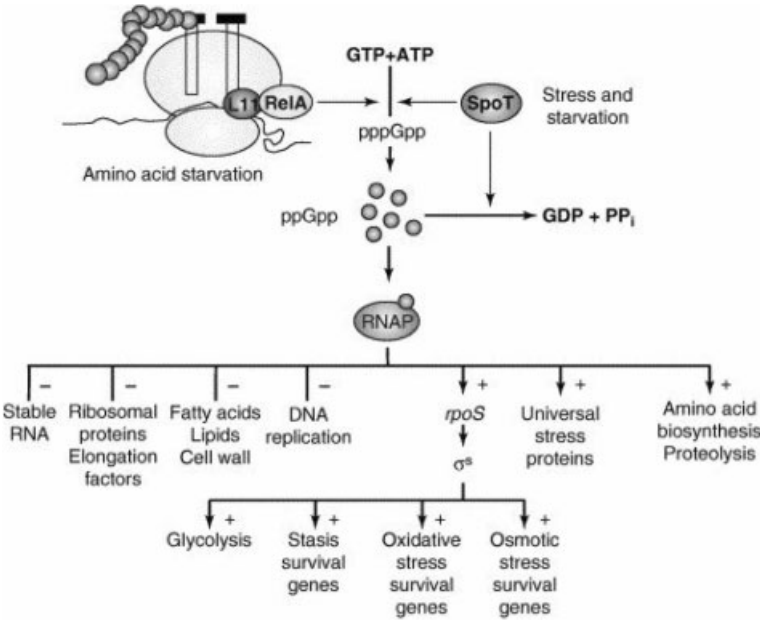


Fig. 9.15 The stringent response. During amino acid starvation, the ribosome-associated protein RelA or the soluble protein SpoT (activated after different stresses and general starvation) produce first pppGpp out of GTP and ATP. This is then further converted to ppGpp by either RelA or SpoT, where SpoT can also cleave ppGpp to GDP and PPi. ppGpp interacts with the RNA polymerase, which in turn stops transcribing genes involved in different pathways denoted by a "minus" and actively transcribes genes marked by a "plus" in the schematic representation. L.U. Magnusson, et al. **2005**, *Trends Microbiol.* 13, 236–242; Fig. 1.

polymerase and decreases open complex stability, accentuating the negative effects of ppGpp on rRNA promoters. The positive effect on the regulation of *rpoS* appears to occur at the translational level.

L.U. Magnusson et al. **2005**, ppGpp: a global regulator in *Escherichia coli*, *Trends Microbiol.* 13, 236–242.

9.8 Nutrient Limitations (Starvation Stress)

During the exponential growth phase in the absence of any stress, σ^S is hardly detectable. This alternative sigma factor rapidly accumulates in response to starvation for sources of nitrogen, carbon, phosphorous or for amino acids, shift to high osmolarity, to high and low temperature (42 °C or below 30 °C, respectively), to acid pH, stationary phase (hence its designation: σ^S for sigma of the stationary phase) and perhaps high cell density. σ^S, encoded by the *rpoS* gene, begins to ac-

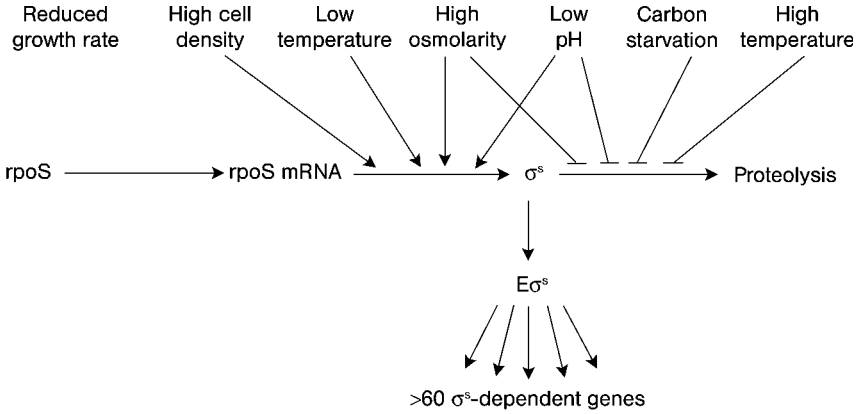


Fig. 9.16 Regulation of active σ^S and factors influencing this regulation. Synthesis of active σ^S is regulated at the level of transcription, translation and stability. Factors influencing one of these three steps positively are indicated by an arrow (e.g., reduced growth rate stimulates transcription of the *rpoS* gene), while factors inhibiting proteolysis are indicated by a T bar. R. Hengge-Aronis **2002**, *Microbiol. Mol. Biol. Rev.* 66, 373–395; Fig. 1.

cumulate in the cell at late exponential growth phase and becomes predominant during the stationary growth phase. It associates with the core enzyme and directs the transcription of genes essential for stationary phase survival. Regulation of σ^S occurs at the transcriptional and post-transcriptional level and involves numerous regulators (Fig. 9.16). σ^S accumulates at the beginning of the stationary phase because many factors act in concert:

- increased transcription of the *rpoS* gene
- translation of the *rpoS* mRNA
- enhanced protein stability
- σ^S activity.

Increased Transcription of *rpoS*

The *rpoS* gene is transcribed from three different promoters, where two precede the bicistronic *nlpD-rpoS* operon and the third is located within *nlpD*. The major transcript originates at this internal promoter, which accumulates when cells growing in rich medium enter the transition phase (end of the exponential and beginning of the stationary phase). cAMP-CRP acts as a negative regulator of *rpoS*; and a putative cAMP-CRP binding site is present upstream of the internal promoter. Three small molecules have been identified to positively influence transcription of *rpoS*: namely ppGpp, homoserine lactone and polyphosphate. While ppGpp positively affects transcriptional activation, the role of homoserine lactone and polyphosphate remains elusive.

Translation of the *rpoS* Transcript

The translation initiation region of the *rpoS* mRNA includes the Shine–Dalgarno sequence, the initiation codon and a putative downstream box which is assumed to form a secondary structure in which the largest part of the Shine–Dalgarno sequence would be base-paired and therefore not accessible for ribosomes. In addition, the nucleoid-associated protein HU, the histone-like protein H-NS and the small regulatory RNAs DsrA and RprA in conjugation with the RNA chaperone Hfq control translation of *rpoS* by modulating the secondary structure of the *rpoS* mRNA. Furthermore, two small regulatory RNAs play a role in *rpoS* translational control. One is the DsrA RNA, which is induced upon a cold shock (Fig. 9.17) and the second, OxyS, is induced in response to oxidative stress (H_2O_2).

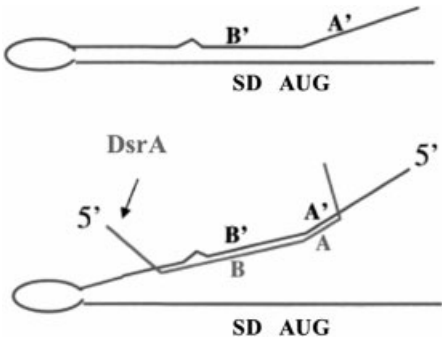


Fig. 9.17 Control of translation initiation of the *rpoS* transcript by the small ncRNA DsrA. While under physiological conditions, access to the Shine–Dalgarno sequence is impaired by the formation of a secondary structure, a temperature downshift induces synthesis of the small ncRNA DsrA, which anneals to the 5′ UTR of the *rpoS* transcript to allow its translation. S. Gottesman **2002**, *Genes Dev.* 16, 2829–2842; Fig. 3, modified.

Control of σ^S Stability

In exponentially growing cells, σ^S can be detected, but the actual levels are very low because of continuous proteolysis. Under these conditions, the half-life of σ^S has been measured to be between 1 min and 4 min. However, its half-life is rapidly stabilized when cells experience carbon starvation or any other stressful situation mentioned above. The protease responsible for the degradation of σ^S is ClpXP. But this protein needs the targeting or adaptor protein RssB, which directly binds to σ^S. RssB seems to be a specific response regulator, which is phosphorylated at aspartate 58 of its receiver domain; and RssB~P targets σ^S to the ClpXP protease. Appropriate stress factors lead to the dephosphorylation of RssB~P, resulting in stabilization of σ^S. Proteolysis of σ^S also depends on a *cis*-acting element called *turnover element*. The core of this turnover element involves the amino acid residues K173, E174 and V177, where K173 is crucial for recognition by RssB~P. K173 is part of region 2.5 involved in transcription initiation, recognizing an extended –10 region as described for σ^{70} (see Section 6.2.2).

Regulation of σ^S Activity

The total concentration of σ^S does not exceed one-third of the concentration of σ^{70}, even at the onset of stationary phase when the σ^S concentration is at its high-

est. In addition, among all sigma factors, σ^S has the lowest affinity for the core enzyme. Several factors have been identified modulating the activity of σ^S. ppGpp modulates competition in favor of σ^S; and glutamate, trehalose and inorganic polyphosphate modulate the activity of σ^S holoenzyme at the steps of holoenzyme formation and/or holoenzyme binding to promoters. In addition, a protein has been identified in directing $E\sigma^S$ to at least one specific promoter. This protein, Crl (from *curli*), is produced at the onset of stationary phase, interacts directly with σ^S and strengthens $E\sigma^S$ binding to the target promoter preceding the *csgAB* operon (codes for curli, see Section 1.6). But there is also a negative regulator of σ^S activity, RssB, which can act like an anti-sigma factor. In addition to increasing the amount of active σ^S, the amount of active σ^{70} is decreased by two different mechanisms during the stationary phase. First, the Rsd protein is produced, acting as an anti-sigma factor of σ^{70}; and second, the small ncRNA 6S RNA is synthesized, which specifically sequesters $E\sigma^{70}$.

R. Hengge-Aronis **2002**, *Microbiol. Mol. Biol. Rev.* 66, 373–395.

9.9 Envelope Stress Response

E. coli is able to adapt to a variety of stresses affecting its envelope (inner membrane, periplasm, outer membrane), causing extracytoplasmic protein misfolding, using three different signal transduction pathways. These consist of the two-component systems CpxAR and BaeSR and the ECF sigma factor σ^E (Fig. 9.18) While some inducing cues are common to two or even all three pathways, each pathway also recognizes unique signals and appears to have distinct physiological roles in the cell.

The Cpx Pathway

The Cpx pathway is controled by the sensor kinase CpxA and the response regulator CpxR. CpxA consists of a periplasmic domain and a conserved cytoplasmic signaling domain separated by two transmembrane α-helices. CpxA senses a wide variety of envelope pertubations (alkaline pH, overexpression of the lipoprotein NlpE, misfolded pilus subunits, alterations in membrane composition), all of which are predicted to result in protein misfolding. While the mechanism of induction remains elusive, activation of CpxA most probably results in relief of an inhibitory interaction between CpxP and the periplasmic sensing domain of CpxA through titration by denatured proteins in the periplasm (Fig. 9.19). Binding of CpxP, a small protein being member of the Cpx regulon, to CpxA is thought to keep it in its phosphatase state to dephosphorylate any CpxR~P present. In addition, CpxP negatively regulates expression of the genes of the Cpx regulon by interacting with the periplasmic domain of CpxA, keeping the pathway off. Upon dissociation of CpxP, CpxA autophosphorylates and then transfers this phosphate to a conserved aspartate in the N-terminal domain of CpxR. CpxR~P in turn func-

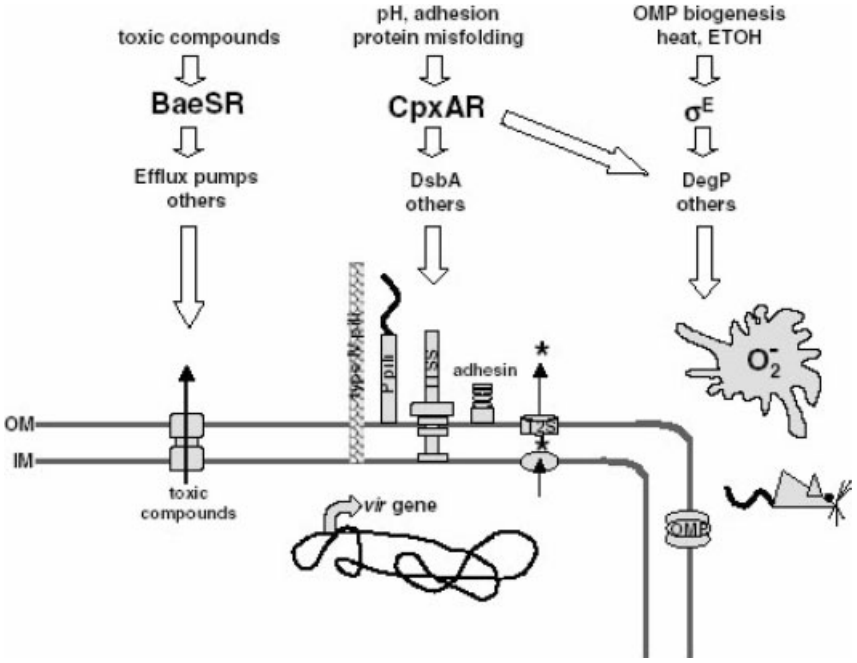


Fig. 9.18 The three envelope stress pathways present in *E. coli* and in Gram-negative bacterial pathogens. The BaeSR two-component system responds to antimicrobial compounds, while the CpxAR two-component system senses denatured proteins present in the periplasm including pilins and adhesins. The σ^E regulon has been shown to become activated by a subset of outer membrane proteins. T.L. Raivio **2005**, *Mol. Microbiol.* 56, 1119–1128.

tions as a transcription factor to activate the expression of more than 100 target genes. A subset of these target genes are involved in envelope folding and degradation functions, such as the periplasmic protease/chaperone DegP (also called HtrA), the major disulfide oxidase DsbA and the two peptidyl-prolyl-isomerases PpiA and PpiD. Another member of the Cpx regulon is the *spy* gene with a so far unknown function. Induction of *spy* expression by severe envelope stress such as spheroblasting is only partially dependent on an intact Cpx pathway (see below). In addition, transcription of some target genes is repressed. Based on these findings, a major role of the Cpx response appears to be to maintain the envelope protein-folding status in the presence of stressful conditions. But in addition, the Cpx pathway is involved in sensing and fostering adhesion and plays an important role in adhesion to abiotic, hydrophobic surfaces through the outer membrane lipoprotein NlpE and in monitoring and altering production of pili.

The Cpx system seems also to be involved in the regulation of early steps in infection, including adherence and possibly invasion, in several pathogens. Cpx senses misfolded pilus subunits in the periplasm during assembly and activates the expression of at least two genes required for efficient P pilus assembly, namely the genes *degP* and *dsbA*. In addition, this system is also involved in the

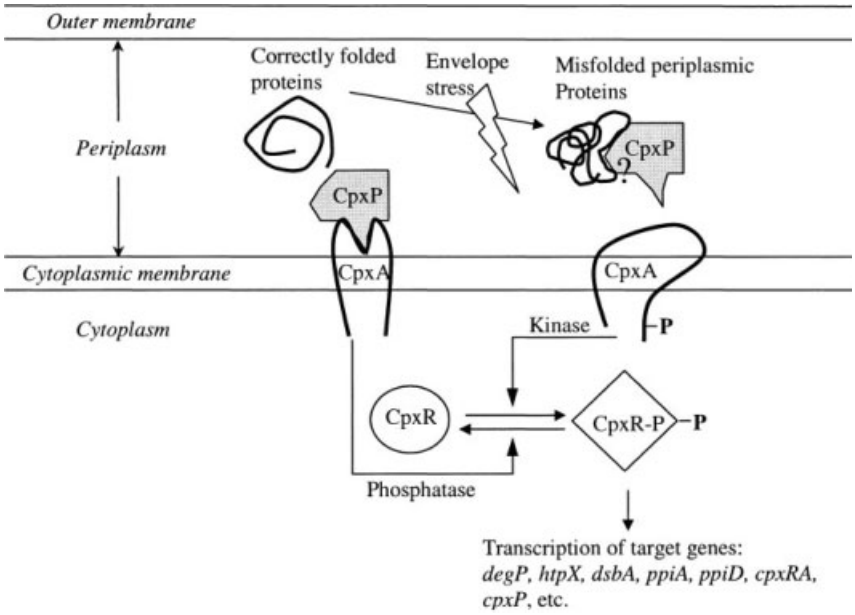


Fig. 9.19 The Cpx signaling system responds to denatured proteins in the periplasm. In the absence of denatured proteins, the CpxA sensor kinase is kept in its phosphatase form through interaction with CpxE. The appearance of denatured proteins causes CpxP to dissociate and bind to these non-native proteins. This causes a conformational change of CpxA, now able to autophosphorylate and transfer the phosphoryl group to the response regulator CpxR. CpxR~P acts as a transcriptional activator for the genes of the Cpx regulon. Upon removal of the denatured proteins, CpxE binds again to CpxA, to convert it back into a phosphatase. The phosphatase activity dephosphorylates CpxR~P, abolishing transcription of the genes of the σ^E regulon. L.M. Wick, T. Egli **2004**, *Adv. Biochem. Eng. Biotechnol.* 89, 1–45; Fig. 2.

expression of the type IV bundle-forming pilus of enteropathogenic *E. coli* strains, an adhesive organelle with a structure and assembly pathway different from that of the P pilus. These observations suggest that the Cpx system is involved in the correct assembly of pilus organelles with diverse structures and assembly pathways.

The BaeSR Pathway

The BaeS/BaeR two-component system controls a second envelope stress response in *E. coli* that induces expression of a distinct set of genes, with the exception of *spy*, which is also a target of the Cpx signaling pathway. The BaeSR signaling system is involved in resistance to several classes of antimicrobial compounds, including bile salts, novobiocin and β-lactam antibiotics through the regulation of genes encoding multidrug efflux pumps.

The σ^E Pathway

The σ^E response pathway is specifically activated by denatured proteins appearing in the periplasm, e.g., after a severe heat shock or overproduction. At least two different pathways lead to the activation of σ^E, one involving a subset of outer membrane proteins with the C-terminal YQF motif and the other so far not studied periplasmic and outer membrane proteins without the motif. Signal transduction is mediated by an elegant network of proteolytic cleavages of the membrane-bound anti-sigma factor RseA (*r*egulation of *s*igma *E A*) by three different proteases acting consecutively.

RseA is a bitopic cytoplasmic membrane protein. The N-terminus of RseA is exposed in the cytoplasm where it binds σ^E, while its C-terminus is located in the periplasm. When denatured proteins appear in the periplasm, either after an extreme heat shock to 45 °C and more, or by overproduction of an outer membrane protein, σ^E is released and activates expression of 54 genes, including those involved in outer membrane protein and outer membrane biogenesis, the DegP protease/chaperone, the petidyl-prolyl-isomerases FkpA and SurA and the periplasmic chaperone Skp. The underlying mechanism has been worked out in the case of outer membrane proteins exposing their carboxy-termini ending in the YQF-COOH or YYF-COOH signature; and it involves the sequential action of three different proteases. The protease DegS is anchored in the inner membrane and its proteolytic site is exposed into the periplasm (Fig. 9.20). C. Gross and coworkers suggested the following model. Normally, the proteolytic site of DegS is inhibited by its PDZ domain. Upon exposure of the YQF/YYF signature due to unfolding of the substrate proteins, the PDZ domain binds to the C terminus of these proteins; and the proteolytic site of DegS attacks RseA to cleave it. This causes a conformational change in the remaining part of RseA, making it prone to receive a second cleavage within or near the membrane-spanning segment, causing release of the N-terminal part into the cytoplasm with σ^E still bound. Next, the new C terminus of the truncated RseA is recognized by the adaptor protein SspB which targets the RseA-σ^E complex to the ATP-dependent ClpXP protease. The ClpX ATPase subunit unfolds RseA and feeds it into the proteolytic chamber of ClpP where it is cleaved into peptides which are released into the cytoplasm. This last proteolytic step finally results in the release of σ^E now able to bind to the RNA polymerase core enzyme.

σ^E mutants have dramatic effects on the virulence of pathogenic bacteria. For instance, *M. tuberculosis rpoE* mutants exhibit decreased survival in macrophages, suggesting diminished capacity to withstand intracellular killing. Activation of a σ^E-like response in this and other species provides a capacity to survive in and withstand the deadly defenses of the host environment. The σ^E and Cpx responses seem to play distinct roles in infection in these pathogens. While the Cpx response is needed early in infection to guarantee successful adhesion and the expression of invasion determinants, the σ^E response appears to be involved in survival within the pathogen.

T.L. Raivio **2005**, Envelope stress responses and Gram-negative bacterial pathogenesis, *Mol. Microbiol.* 56, 1119–1128.

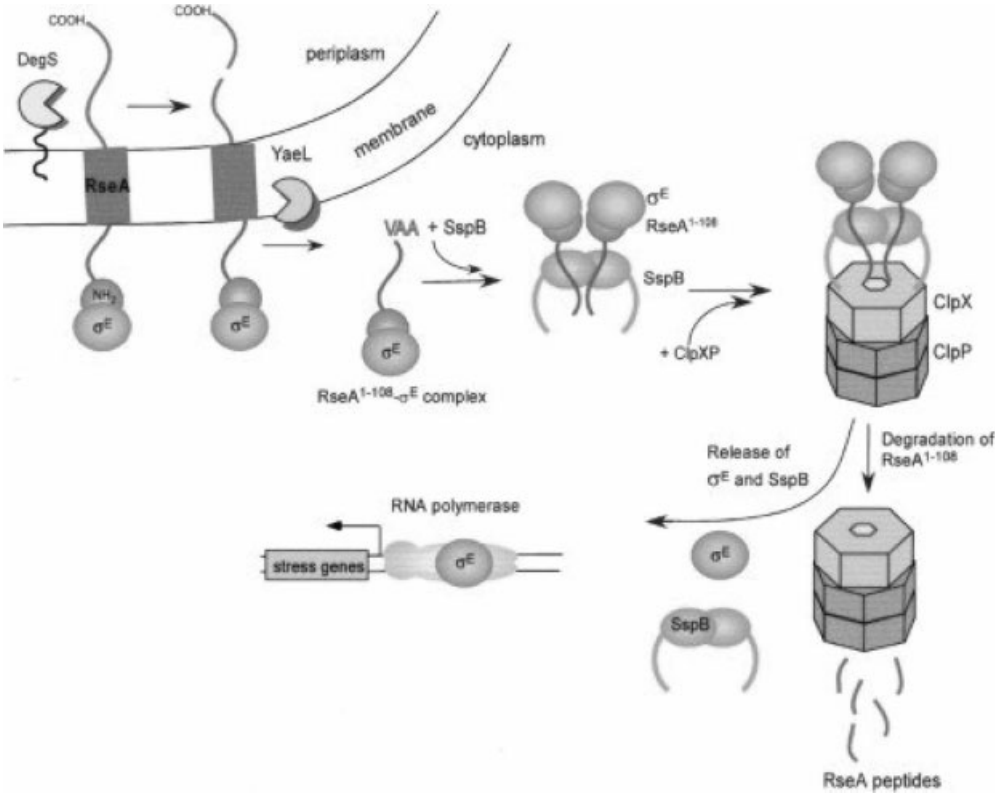


Fig. 9.20 Activation of σ^E needs the successive action of three different proteases. One way to release σ^E from its anti-sigma factor RseA involves the appearance of a subset of outer membrane proteins carrying the YQF signature at their immediate C terminus. In its unfolded form, this signature is recognized by the PDZ domain of the membrane-anchored DegS protease. This in turn leads to activation of the proteolytic activity of DegS, cleaving RseA in its periplasmic domain. Next, the RseP protease cuts within or near the transmembrane segment, causing release of the N-terminal part of RseA with σ^E still bound. The C terminus of the truncated RseA is recognized by the SspB adaptor protein targeting RseA to the ClpXP protease, where it is unfolded and degraded into peptides. The last proteolytic step finally leads to the release of σ^E. B.M. Alba, C.A. Gross **2004**, *Mol. Microbiol.* 52, 613–619. J.M. Flynn, et al. **2004**, *Genes Dev.* 18, 2292–2301; Fig. 4. (This figure also appears with the color plates.)

10
Exchange of Genes: Transformation, Conjugation and Transduction

Bacteria can acquire genes by three different mechanisms, collectively called *h*orizontal *g*ene *t*ransfer (HGT): transformation, transduction and conjugation. *Transformation* involves the uptake of free DNA from the extracellular environment, and *genetic competence* is the ability to undergo transformation. During *transduction*, the DNA is transferred using bacteriophages as carrier. While some phages can transfer any gene with the same probability (so-called generalized transducing phages), other pick up only specific genes (specialized transducing phages). During *conjugation*, genes are directly transferred from one bacterial cell to another. A fourth mechanism can be envisaged though there is no experimental proof. Since some bacteria release proteins in membrane vesicles and others secrete DNA, there is the theoretical possibility at least that DNA fragments are released embedded in membrane vesicles which can fuse with the cytoplasmic membrane of Gram-positive bacteria to enter their cytoplasm. Why did bacteria develop DNA uptake systems? Three models can account for the evolution of uptake systems:

- DNA increases the genetic diversity: the acquisition of useful genetic information such as antibiotic resistance genes, novel metabolic pathways or virulence genes.
- DNA repair: DNA from the same or closely related species can serve as template for the repair of DNA damage.
- DNA as nutrient: DNA can be used as a source of carbon, nitrogen and phosphorous.

10.1
Transformation

Transformation refers to the uptake of free DNA by a bacterial cell and its stable integration into the bacterial genome. Approximately 90 transformable bacterial species have been described so far, but not all are known to be competent for DNA uptake in the natural environment, where free DNA is abundant as it is released from dead organisms. These include archaea and divergent subdivisions of bacteria, among them representatives of the Gram-positive bacteria, cyanobacter-

Dynamics of the Bacterial Chromosome. Structure and Function. W. Schumann
Copyright © 2006 Wiley-VCH Verlag GmbH & Co. KGaA
ISBN 3-527-30496-7

ia, *Thermus* spp, *Deinococcus* spp, green sulfur bacteria and many other Gram-negative bacteria. Many human pathogenic bacteria, including representatives of the genera *Haemophilus*, *Helicobacter*, *Staphylococcus*, *Streptococcus* and *Campylobacter* are naturally transformable, suggesting that transformation is functionally important in the environment, giving access to DNA as a source of nutrients or genetic information. In addition, DNA can also be actively secreted by viable organisms such as some strains of *N. gonorrhoeae* that use a type IV secretion system to release DNA into the environment.

Transformation requires the induction of genetic *competence*, a physiological state that allows cells to take up DNA, which has been found to involve approximately 20–50 proteins. The different steps in transformantion involve DNA binding, DNA fragmentation, uptake of DNA fragments and stable maintenance of the acquired DNA either by recombination or by recircularization in the case of plasmids. For the stable recircularization of plasmid DNA, two single strands with partially complementary sequences must enter the cell to allow efficient repair by host enzymes. In liquid culture, development of competence is growth-phase dependent in most organisms studied so far. In contrast, biofilms of *Actinetobacter* spp grown in flow cells are permanently competent and other bacterial species are always capable of taking up DNA, such as *T. thermophilus* and *H. pylori*. The uptake mechanisms are similar in Gram-positives and Gram-negatives and include type IV pili (syringe-like machinery able to transport protein or DNA effector molecules directly into eukaryotic cells; see Section 8.6.4) and a type II protein export system (allows the energy-dependent secretion of specific proteins from the periplasm of Gram-negative bacteria; see Section 8.6.2). Only a few species, including *B. subtilis*, *S. pneumoniae* and *Acinetobacter* spp, are able to take up any DNA from the environment. Most species recognize related DNA by specific nucleotide sequence tags overrepresented in their own genomic DNA. It is 9–11 bp in length and present once per 4–5 kb.

Release of Extracellular DNA in the Environment

DNA continually enters the environment upon release from decomposing cells, disrupted cells or viruses or through excretion from living cells. The stability of intact DNA in the environment depends on the activity and location of nucleases and reactive chemicals. Active excretion of DNA has been reported for many genera of bacteria, as already mentioned, including *Acinetobacter* (1–3 μg ml^{-1} in liquid cultures), *Alcaligenes*, *Azotobacter*, *Bacillus*, *Flavobacterium*, *Micrococcus*, *Pseudomonas* (up to 780 μg) and *Streptococcus*. The role of release of DNA by bacteria in natural habitats remains to be fully understood. Passive release of DNA from dead bacteria occurs after self-induced lysis, while pathogenic bacteria can undergo lysis caused either by the host immune system (between 95% and 100% of the bacterial DNA is released after contact with the immune system, but most of this DNA is destroyed by DNases present in human serum and plasma) or the antibiotic treatment of infections. DNA can persist in the blood long enough for natural transformation of infectious bacteria to occur. It has been shown that DNA re-

leased by pathogenic and commensal bacteria present in the intestine can temporarily enter the bloodstream and reach various organs in mammals.

Stability of Free DNA in the Environment

The persistence of DNA will be determined by the environmental conditions. In soil and sediments, approximately 1 μg DNA can be recovered per gram of material, while in fresh and marine water 0.003–88.0 μg dissolved DNA can be found per liter. Adding purified DNA into soil, water, silage or various types of food has shown that it is not immediately degraded and is present for hours to days, as measured in natural transformation assays.

Uptake of DNA into Bacteria

Upon exposure to competent cells, the DNA binds noncovalently to sites present on the cell surface. The number of binding sites ranges from 30 to 80 for those species where it has been determined. The next step, uptake into the cytoplasm, varies among bacterial species and has been characterized for only a few species so far (see below). The double-stranded DNA is converted into single-stranded DNA during translocation across the inner membrane. The uptake of plasmids requires complex steps to reassemble duplex molecules in the recipient cytoplasm. This requires either the presence of the plasmid in the recipient cell already, or for the plasmid to be multimeric or to enter in multiple copies. The uptake of DNA has been measured for *S. pneumoniae* and occurs at about 100 bp s^{-1}. The internalized DNA normally persists only transiently in the cytoplasm owing to the inability of the DNA fragments to replicate and due to degradation by nucleases. If single-stranded DNA fragments assembled into double-stranded DNA, they might be attacked by restriction enzymes, causing their fragmentation.

Recombination of DNA with the Recipient Genome

DNA taken up by transformation must integrate into the bacterial genome to persist for many generations. Three mechanisms have been described: homologous recombination, illegitimate recombination and additive integration. For homologous recombination, incoming DNA must contain regions 25–200 bp in length of high similarity to the recipient genome. These regions will initiate DNA pairing and strand exchange. It could be measured that *B. subtilis* and *S. pneumoniae* recombine 25–50% of the internalized DNA fragments. A log-linear decrease in recombination frequencies with increasing divergence has been found with several bacterial species. The need for sequence similarity for integration may be circumvented by illegitimate recombination. In *E. coli*, several genes have been identified as controling illegitimate recombination through double-stranded breaks and end-joining through illegitimate recombination events occur at a very low frequency (see Section 4.3.6). Additive integration occurs between two circular molecules where at least one is able to replicate. This process is based on single crossovers at

short regions with high DNA sequence similarity and leads either to the fusion of plasmids or the integration of circular DNA molecules into the bacterial chromosome. Experiments carried out in *E. coli* have shown that homologous sequences of at least 25 bp are sufficient to detect recombination.

10.1.1 Transformation of Gram-positive Bacteria

Bacillus subtilis

In order to cope with the strong nutritional and physical fluctuations related to its natural biotope, *B. subtilis* developed a wide arsenal of survival strategies. At the end of the exponential growth phase, when nutrients become limiting for optimal growth, cells synthesize a complex motility and chemotaxis system which, in their natural habitat, enable them to search for nutrients. If nutritional limitation continues, the motile cells enter the stationary growth phase and start to secrete degradative enzymes such as proteases to degrade proteins and to produce antibiotics to kill potential competitors. Prolonged nutritional stress results in the development of competence and ultimately in the formation of endospores.

Genetic competence converts *B. subtilis* cells into transformable cells. There is no nucleotide sequence specificity for DNA binding and DNA uptake; and competent cells can incorporate phage, plasmid and chromosomal DNA. To take up DNA from the medium, cells synthesize a specific DNA binding and uptake system, the competence machine, and there are about 50 competence machines per cell. The competence machine (Fig. 10.1) consists of a pseudopilus consisting of the major pseudopilin (ComGC) and three minor pseudopilins (ComGD, ComGE, ComGG) which are synthesized as prepilins and then processed by the prepilin peptidase (ComC). The polytopic membrane protein ComGB and the traffic NTPase ComGA assist formation of the pseudopilus which traverses the thick cell wall. Though the pseudopilus does not bind DNA, it is assumed to guide it to the receptor protein ComEA; and an integral membrane protein with its C-terminal portion located outside of the membrane is the DNA receptor for transformation. Its DNA-binding activity was localized to its C-terminal domain, which contains a helix–hairpin–helix motif, predicted to bind DNA without sequence specificity. ComEA also seems to participate in DNA transport involving a flexible stretch in the protein (QQGGGG), located immediately before the DNA-binding domain, which could deliver the DNA to the entrance of the channel on the cytoplasmic membrane.

After binding to ComEA, donor DNA undergoes limited double-stranded breakage catalyzed by NucA, another integral membrane protein with endonuclease activity. Fragments of a length between 15 kb and 20 kb are produced where the newly formed ends generated by the double-strand cleavage are probably the preferred initiation points for transport. Next, one strand is degraded to acid-soluble products which are released into the medium. The single-stranded DNA molecule is transported through the cytoplasmic membrane, which requires the putative channel proteins ComEC (model: two ComEC proteins form a channel) and the

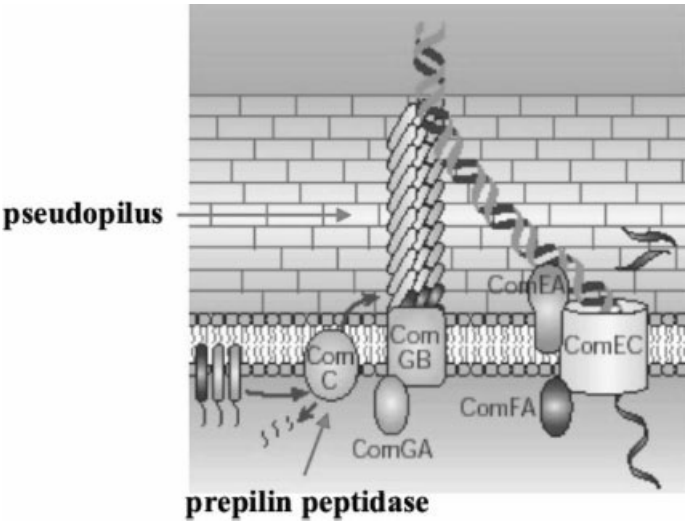


Fig. 10.1 The *B. subtilis* competence machine. The pseudopilus consists of the major pseudopilin (ComGC) and three minor pseudopilins (ComGD, ComGE, ComGG) which are processed by the prepilin peptidase (ComC). The polytopic membrane protein ComGB and the traffic NTPase ComGA assist this process. DNA molecules first make contact with the pseudopilin, then bind to the receptor protein, ComEA, which delivers the DNA to the ComEC channel located in the cytoplasmic membrane. Only one strand enters the cytoplasm; and the ATP-binding protein ComFA is involved in the transport process. I. Chen, D. Dubnau **2004**, *Nat. Rev. Microbiol.* 2, 241; Fig. 1, modified. (This figure also appears with the color plates.)

ATPase ComFA. The DNA crosses the membrane in linear fashion without observable pausing events; and proton motive force acts as a major energy source. An uptake rate of 80±10 nucleotides s^{-1} have been measured. Incoming single-stranded DNA presumably associates with the homologous recombination protein RecA and, together with the other recombination proteins, catalyzes the integration of the newly acquired DNA into the chromosome. Plasmids can be reconstituted, followed by their replication.

The expression of the genes of the competence machine is regulated by two different modules. Quorum sensing and nutritional limitation activate module 1, resulting in the synthesis of the small protein ComS. ComS, in turn, activates module 2 the output of which is the competence transcription factor ComK. A sufficiently high cell density is a prerequisite for competence to develop optimally. In addition, medium constituents are important regulatory factors. Cell density is measured by two different quorum-sensing pheromones (see Section 6.2.8.6). Accumulation of these pheromones in the growth medium signals the presence of a sufficient number of cells. The main competence stimulating factor, ComX, is an oligopeptide of 9–10 amino acids, with an isoprenyl modification at a tryptophane residue. ComX is sensed by the sensor kinase ComP, which undergoes autophosphorylation and subsequent transfer of the phosphoryl group to the response

regulator ComA. The second quorum-sensing pathway converges with the ComX pathway at the point of ComA phosphorylation. ComA~P stimulates transcription of the bicistronic *srfA* operon. This operon codes for a very large protein complex responsible for the nonribosomal synthesis of the lipopeptide antibiotic surfactin. Surfactin is a potent biosurfactant which is able to lyse microorganisms, causing release of their DNA which can be taken up by competent cells. The second open reading frame encodes a 46-amino-acid peptide, designated ComS, that is essential for genetic competence to develop. The second pheromone is the *c*ompetence-*s*timulating *f*actor, CSF. CSF is encoded by the *phrC* gene and synthesized as a 40-amino-acid peptide which is secreted into the medium, processed outside and taken up as a pentapeptide (ERGMT) by the oligopeptide permease Spo0K. The target of CSF is the RapC protein, which binds to ComA independently of its phosphorylation state and inhibits its DNA-binding activity. RapC contains six tetratricopeptide (TRP) repeats which mediate interaction with the pentapeptide inhibitor PhrC or with its target protein ComA. TRP repeats are structural domains found in a wide variety of proteins from pro- to eukaryotes. Each repeat consists of 34 amino acids that fold into a structure consisting of a pair of antiparallel α-helices of equivalent length generally associated with a packing angle of ~24° between the helix axes. TRP motifs are usually found in tandem arrays ranging between three and 16 motifs per protein. Therefore, a low CSF concentration leads to the dephosphorylation of ComA~P while a high concentration favors stability of ComA~P. The second module regulates the activity of the transcription factor ComK. ComK activates its own transcription by binding to its own promoter, where two binding sites are present. These two binding sites are 31 bp away; and efficient binding is aided by DegU interacting with the DNA in between these two ComK-binding sites. Once sufficient amounts of ComK are present, DegU is not needed any more. To prevent inappropriate competence, the small amount of ComK produced during exponential growth is bound by the adapter protein MecA (see Section 7.4.1) and targeted to the ClpCP protease where ComK is degraded. How is degradation of ComK prevented in those cells committed to develop competence? When ComS has been produced, it binds to MecA and stimulates the degradation of both proteins by the ClpCP complex. In summary, the two competence pheromones ComX and PhrC (CSF) stimulate *comS* transcription which in turn controls *comK* expression on the post-translational level. It should be mentioned that, even under optimal conditions, only about 10–20% of the cells in a culture will produce ComK.

Streptococcus pneumoniae

DNA uptake starts also with its binding to the cell surface without any sequence specificity; and 33–75 uptake sites per cell have been determined. DNA uptake is followed by fragmentation, where single-strand nicks occur first and double-strand breaks subsequently. The average length between nicks is 6 kb. Transport of DNA is similar to that of *B. subtilis* in most aspects. It is assumed that transport initiates at a second break opposite the initial single-stranded nick and one strand

is taken up in the 3′→5′ polarity at an uptake rate of 90–100 nucleotides at 31 °C, while the second strand is degraded with 5′→3′ polarity. A ComEA receptor protein ortholog has been detected; and the EndA endonuclease is required for transport and degradation of the nontransported strand, but not to introduce the initial nick. Proteins involved in formation of the competence pseudopilus and the DNA translocation machinery similar to those of *B. subtilis* have been identified, too.

The regulation of competence is known to rely on a cell–cell signaling system involving five genes: *comABCDE*. *comC* encodes the preCSP (*c*ompetence-stimulating *p*eptide), which is matured and exported by ComAB as an unmodified 17-residue-long peptide pheromone. ComDE is a two-component regulatory system that responds to external CSP. ComD, a membrane-bound histidine kinase, is the CSP receptor, which upon signal recognition activates its cognate response regulator, ComE. ComE~P activates transcription of the early *com* genes, among them *comX* coding for the alternative sigma factor σ^H. Eo^H is responsible for transcription of the late *com* genes. The entire *com* regulon consists of more than 100 genes.

10.1.2 Transformation of Gram-negative Bacteria

N. gonorrhoeae

N. gonorrhoeae serves as the prototype for the group of Gram-negative bacteria. In contrast to *B. subtilis*, DNA uptake by cells of this species and others is sequence-specific. While *Neisseria* sp. recognizes the sequence 5′-GGCGTCTGAA-3′, *H. influenzae* binds specifically DNA fragments containing 5′-AAGTGCGGT-3′ (and its inverse complement ACCGCACTT). The latter sequence is also recognized by the related bacterium *A. actinomycetemcomitans*. As the genomes of these bacterial species are rich in these uptake sequences (a total of 1471 copies in the chromosome of *H. influenzae* strain Rd; 1891 copies in that of *N. meningitis*), uptake of homospecific sequences is favored. Specific receptor proteins have not yet been identified; and in *N. gonorrhoeae* specific binding of DNA depends on the presence of the major pilin. The potential receptor recognizes the DNA uptake sequence and triggers the uptake process, involving the transport of DNA across the outer membrane into either the periplasmic space or specialized vesicle structures.

Crossing the outer membrane involves *secretins*. Secretins are outer-membrane proteins that form stable multimeric donut-like structures, with 12 or 14 subunits and a central cavity ranging from 6.0 nm to 8.8 nm (Fig. 10.2). The best characterized secretin is PilQ from *N. gonorrhoeae*, which also functions in pilus biogenesis where it is required to extrude the pilus filament across the outer membrane. The central cavity in the PilQ 12-mer forming an aqueous channel, with a diameter of 6.5 nm, could easily accommodate the DNA double helix (~2.4 nm). The putative DNA receptor could participate in signaling the opening of the PilQ channel during transformation, but experimental evidence that the DNA really passes the secretin-formed channel is still lacking. In addition to the secretin channel, DNA uptake needs a structure called pseudopilin. The major pilin PilE

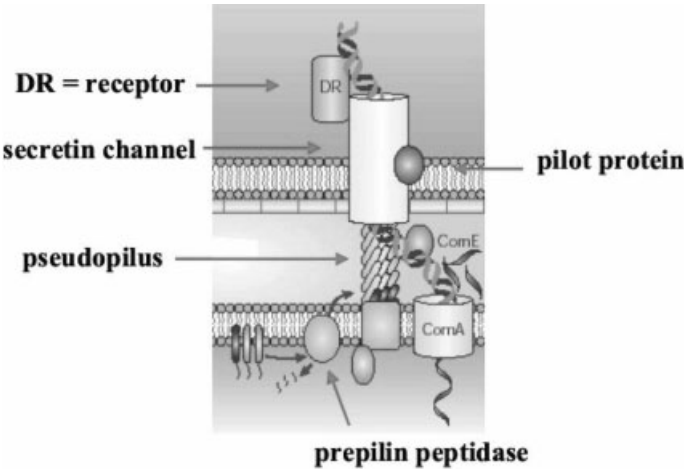


Fig. 10.2 The competence machine of *N. gonorrhoeae*. This machine consists of a secretin forming a ring-like structure in the outer membrane which is connected to the pseudopilin (PilE, ComP), the subunits of which are processed by the prepilin peptidase (PilD). DNA molecules are recognized by the hypothetic receptor protein (DR) and delivered to the secretin (PilQ). Transport occurs through the water-filled secretin channel with the assistance of the protein PilP. The incoming DNA is recognized by a periplasmic DNA-binding protein (ComE) and delivered to a channel in the inner membrane (ComA). Only one strand enters the cytoplasm; and the other is degraded in the periplasm. I. Chen, D. Dubnau **2004**, *Nat. Rev. Microbiol.* 2, 241; Fig. 1, modified. (This figure also appears with the color plates.)

and the minor pilin PilF are processed by the prepilin peptidase PilD and assembled into the pseudopilin.

The incoming DNA is transported across the outer membrane through the secretin channel with the assistance of its pilot protein PilP. The periplasmic DNA-binding protein ComE is involved in uptake and delivers the DNA to the channel at the cytoplasmic membrane formed by ComA. Only one strand enters the cytoplasm, catalyzed by a still unknown translocase, while the complementary strand is degraded to acid-soluble products.

10.2 Conjugation

Conjugative DNA strand transfer is a highly conserved mechanism for unidirectional transfer of genetic information among bacteria of the same species, from one species to another and, in some instances, between kingdoms. Importantly, this is the mechanism used by many plasmids and a group of mobile elements called conjugative transposons that are derived from phages to facilitate their spread throughout a bacterial population; and the same underlying mechanism is used to transfer T-DNA from *A. tumefaciens* to plant cells. Genetic conjugation was discovered in 1946 by J. Lederberg and E. Tatum. They mixed cultures of two

different auxotrophic *E. coli* K12 strains and obtained prototrophs. Initially, cellular fusion was thought to be responsible for the formation of prototrophs. Later, unidirectional transfer of DNA was demonstrated to be responsible. The existence of the F (*f*ertility) or sex factor responsible for genetic transfer was demonstrated by the observation that this plasmid can be transmitted from F^+ to F^- cells without involving the bacterial chromosome. The choice of *E. coli* K12 was important for this discovery, since this strain harbored a plasmid that was a naturally occurring mutant transferring at elevated frequency. In addition, it contains IS elements, which facilitated its integration into the host chromosome.

Conjugative plasmids encode the proteins needed in their own transfer from a *donor* to a *recipient* cell. The conjugation process can be divided into three steps:

1. The *pili* encoded by the conjugative plasmids from proteobacteria or the *aggregation substances* encoded from the pheromone conjugative plasmids of low G+C Gram-positive bacteria promote cell-to-cell contact.
2. A rolling-circle replication-like system termed relaxosome nicks the transferred strand at the origin of transfer (*oriT*), unwinds the nicked strand, replicates the remaining strand from its free 3′-OH and then joins both ends.
3. The nicked strand is transferred through a mating pore into the recipient cell, followed by its replication to a double-stranded DNA molecule.

Plasmids are classified into *incompatibility* groups. Two different plasmids are designated as incompatible if they do not coexist within the same cell over many generations. Incompatibility is based on shared replication, copy number control or partitioning mechanisms excluding one plasmid from the cell. Plasmids are classified into different incompatibility groups, denoted as Inc followed normally by a capital letter such as IncP, IncW, etc. How can it be tested whether two plasmids belong to the same or to two different incompatibility groups? Let us assume that both plasmids code for different antibiotic resistance marker, e.g., ampicillin and tetracycline resistance. Both plasmids are successively transformed into the same bacterial strain; and the transformed cells are grown in the presence of both antibiotics. Next, the selective pressure is alleviated for 50–100 generations by growth in the complete absence of the two antibiotics, depending on the copy number of the plasmids. Then, the bacteria are plated onto plates containing ampicillin or tetracycline, followed by testing single colonies for growth on the second antibiotic. If the two plasmids are compatible with each other, cells are able to grow in the presence of both antibiotics. If the two plasmids belong to the same incompatibility group, only one will survive, normally the one with the smaller size.

10.2.1 Gram-negative Bacteria

Conjugative systems in Gram-negative bacteria depend on three essential components:

- a transmembrane multiprotein complex called the *transferosome*, which belongs to the type IV secretion system (T4SS) family (see Section 8.6.4), spans the cell envelope and is responsible for the synthesis of the conjugative pilus;
- the *relaxosome*, a complex of proteins that process the DNA at the *ori*gin of *t*ransfer (*oriT*);
- the *coupling protein* connecting the two other entities together.

All conjugative systems code for a pilus and initiate DNA replication and transfer upon contact with a recipient cell, suggesting that the process involves signaling. Most data on the conjugative DNA transfer have been obtained with the 100-kb F-plasmid serving as the paradigm for a large group of conjugative plasmids and integrative, conjugative elements that carry genes important for human and veterinary medicine, such as those for antibiotic resistance and toxin production.

Pilus Structure and Assembly

Pili have a diameter of 6–12 nm and a flexible or rigid structure. Pilus synthesis and maturation has been studied in detail in the case of the F-pilus. This pilus is usually ~2 μm in length and is composed of identical 7.2-kDa pilin subunits arranged in a helical array of five subunits per turn. It is 8 nm in diameter, with an internal lumen of 2 nm, a space wide enough to accommodate a single-stranded DNA molecule as well as a protein covalently attached to its 5′ end in an extended conformation. The F pilin is synthesized as 121-amino acid propilin with a 51-amino acid leader peptide, which is removed by the leader peptidase upon insertion in the inner membrane in a Sec-independent mechanism that requires the 10.9-kDa TraQ, a chaperone-like inner membrane protein. After cleavage, the pilin subunit undergoes acetylation at the N-terminus catalyzed by TraX, where its biological function remains elusive. The first step in F pilus maturation is the formation of a specialized structure at the cell surface, consisting of the unextended pilus tip. Formation of this tip is affected by the following proteins: TraA (pilin), TraL, TraE, TraK, TraC (could form a dimer that assembles into a higher order structure such as a hexameric ring) and TraG. The next step is pilus extension, which needs the proteins TraB, TraF, TraH, TraP (pilus stability), TraW, TraV, TrbI (controls pilus length) and TrbCF. The F pilus serves as receptor for a wide range of bacteriophages, including the small spherical RNA-containing phages such as R17, MS2 and Qβ, the filamentous phages including M13, fd and f1 and a number of double-strand DNA phages. The RNA phages usually attach to the side of the pilus, while the filamentous phages attach to the tip.

The Transferosome

The transferosome encoded by the F plasmid consists of the outer membrane protein TraN, an adhesin which is thought to interact with TraG to form a bridge spanning the donor cell envelope. TraV interacts with TraK that, in turn, interacts

with TraB to form a scaffold for assembling the transferosome. The C-terminal domain of TraK are related to secretins, suggesting that they might form the channel through the outer membrane that is anchored by TraV. TraB is a periplasmic protein with a predicted cytoplasmic membrane anchor and assumed to have an extended structure extending into the periplasmic space. In summary, TraF, TraH, TraU, TraW, TraN and TrbC form a heteromultimeric complex involved in pilus retraction and mating-pair stabilization; and this complex in turn interacts with the TraG-bridging pilus assembly, mating-pair stabilization and pore formation. The coupling protein TraD has two N-terminal membrane-spanning regions, such that the large C-terminal region is in the cytoplasm. Coupling proteins contain at least one consensus nucleotide-binding motif (a Walker A box motif alone or with a Walker B motif). They are thought to use the energy of NTP hydrolysis to couple the activated relaxosome to the transport machinery. Coupling proteins may also serve as a pump to propel the single-stranded DNA through the mating pore.

Mating-pair Formation and Surface Exclusion

In the F conjugative system, donor and recipient cells come into close contact to form *mating junctions*. These junctions have an electron-dense region, probably composed of protein, between the cells. One donor cell can form junctions with several surrounding recipient cells, resulting in *mating aggregates* stable for about 30 min. The junction between mating cells requires that the outer membranes of the two cells come into close apposition. Stabilization of the mating pairs requires two transfer gene products, TraN and TraG, in the donor cell and the LPS inner core and OmpA protein in the recipient cells. TraN (60 kDa) is essential for transfer and recognizes OmpA. TraG is a 103-kDa inner membrane protein with a large periplasmic domain assumed to undergo post-translational processing by the signal peptidase to release a 53-kDa fragment called TraG* into the periplasm.

Most conjugative plasmids encode an exclusion system that reduces redundant transfer between cells harboring the same or closely related plasmids. The F factor encodes two exclusion genes, *traT* and *traS*. The 26-kDa outer membrane lipoprotein reduces stabilization of mating pairs in a process known as surface exclusion. The 16.9-kDa TraS inner membrane protein blocks DNA import from a potential donor cell, termed entry exclusion.

The Relaxosome

Initiation of DNA transfer requires formation of the relaxosome, a nucleoprotein complex consisting of negatively supercoiled *oriT* DNA, a relaxase and accessory proteins (Fig. 10.3). The relaxase introduces a site- and strand-specific break in the phosphodiester backbone at *nic* via a transesterification reaction that leaves the initiator protein covalently bound on the 5′ end of the cleaved DNA strand via phosphotyrosyl linkage. The DNA is subsequently unwound, either through the action of a helicase or as a function of DNA replication, extending from the newly

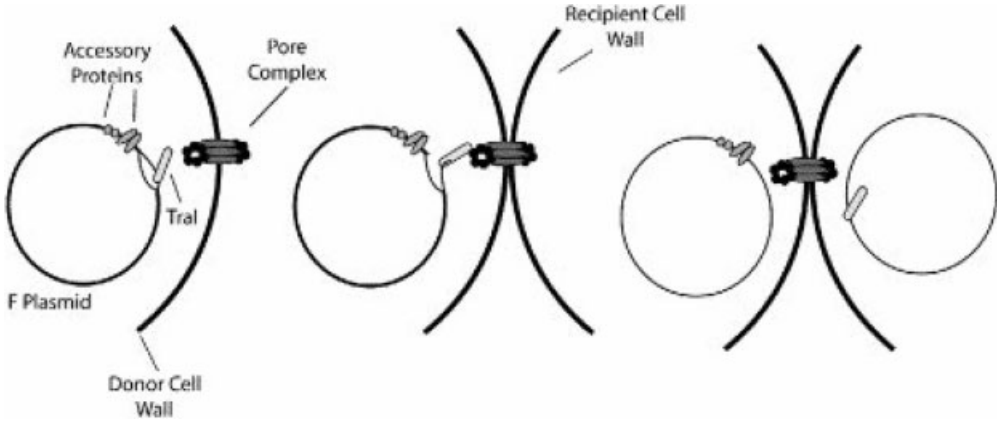


Fig. 10.3 Model for the relaxase function of conjugative plasmids. Accessory proteins bound in the vicinity of the *oriT* facilitate binding of the relaxase to the *oriT*, where it cleaves a specific site, the *nic* site, whereby the relaxase becomes covalently linked to the 5′ end at a nick through a phosphotyrosyl linkage. Next, the cleaved strand is transferred to the recipient cell, where the relaxase recircularizes the single-stranded DNA by reversing the cleavage reaction. C. Larkin, et al. **2005**, *Structure* 13, 1533–1544; Fig. 1.

formed 3′-OH which is used as a primer by the DNA polymerase III via a rolling-circle mechanism to produce the single-stranded DNA molecule that is transferred to the recipient cell. In the case of plasmid DNA transfer, the 3′-OH subsequently acts as a nucleophile to reverse the original transesterification. This restores the integrity of the single-stranded DNA transferred to the recipient and releases the bound relaxase. The action of the relaxase at *oriT* is governed by several auxiliary proteins that act to alter the architecture of the DNA molecule in the vicinity of *nic* to allow binding of the relaxase protein. These auxiliary proteins are both host- and plasmid-encoded in origin. The process of single-stranded DNA transfer proceeds at approximately 45 kb min^{-1} at 37 °C; and F transfer occurs in just 2 min.

The *oriT* region of a plasmid can range from 38 bp to ~500 bp in size and usually contains sequences that dictate intrinsic bends as well as direct or inverted repeats acting as binding sites for the accessory proteins which alter the structure of the DNA and bring the *nic* site in register with the relaxase. In the case of the F factor, the relaxosome is assembled in a stepwise manner, whereby TraY together with IHF binds at *oriT* and directs the TraI relaxase to the *nic* site. The 15.2-kDa TraY, which contains a ribbon–helix–helix motif, binds either as a dimer or monomer form to high-affinity sites near *oriT* and the P_{traY} promoter. TraY is also required for expression of *traM* and the long *traY-traI* operon. The 14.5-kDa TraM forms tetramers and cooperatively binds to three sites (*sbmA, sbmB, sbmC*) at the *oriT*. TraM autoregulates its expression by binding to *sbmA* and *sbmB*, which overlap the two tandem *traM* promoters. TraM interacts with TraD, an inner membrane component of the transferosome. TraM was proposed to transmit a signal between the cytoplasmic relaxosome and the transferosome during conjugation.

Some transfer systems code for a DNA primase gene (*traC* in RP4) that increases the efficiency of conjugation. The TraC protein is transported in large amounts per DNA strand into the cytoplasm of the recipient cell, but not as a nucleoprotein complex rather than by separate transfer. DNA primases are assumed to generate the RNA primers needed for synthesis of the complementary strand.

The single-stranded DNA entering the recipient cell must overcome the hostile environment often present in the new host. The leading region defined as the first portion of a plasmid to enter the recipient cell encodes several proteins ensuring survival. In the case of the F plasmid, the *ssb* and *psiB* genes are highly conserved. While *ssb* encodes a single-stranded DNA-binding protein essential for bacterial DNA replication covering the incoming DNA, *psiB* specifies a protein that prevents induction of the bacterial SOS regulon. Expression of these genes is thought to be the result of transcription from promoters that are found in intramolecular hairpins in the single-stranded DNA. In addition to these two genes, plasmids from several incompatibility groups determine an ArdA antirestriction system protecting incoming plasmids from type I restriction enzymes.

Regulation of Conjugative Transfer

The *tra* region of the F factor contains the three primary promoters P_{traM}, P_{traJ} and P_{traY}, where the latter initiates transcription of the complete *tra* operon from *traY* to *traI*. P_{traY} is positively regulated by TraJ, which interrupts a protein–DNA complex of TraY, SfrA/ArcA and IHF, thereby alleviating repression. In most F-like plasmids, *traJ* is repressed by an antisense RNA, *finP*, in conjunction with an RNA chaperone, FinO. In F, *finO* is interrupted by an IS*3* element, resulting in the derepression of both *traJ* expression and F plasmid transfer. Besides the host factors SfrA/ArcA and IHF and Fis, which have binding sites upstream of P_{traY}, CRP and Lrp control the P_{traJ} promoter. All these proteins presumably act to control F transfer expression in response to environmental signals. SfrA/ArcA is member of a two-component signal transduction system that regulates aerobic respiration functions and fertility via independent regions of the protein. In addition, the CpxA-CpxR two-component regulatory pair (see Section 9.9) has been implicated in regulating *traJ* expression, which is downregulated in a post-transcriptional manner in response to cell envelope stress.

Mobilizable Plasmids

Many small nonconjugative plasmids are mobilizable in the presence of a self-transmissible plasmid. These plasmids code for protein(s) required for recognizing their own *oriT* and cleaving at the *nic* site and transfer of the complex to the transferosome encoded by the coresident conjugative plasmid. The mobilization frequency is dependent on the conjugation system and can vary considerably. For instance, the 6.6-kb ColE1 plasmid, coding for four different *mob* genes, is efficiently mobilized by IncFI, IncI1 and IncP plasmids, but less efficienctly by IncW plasmids. Several important pBR cloning vectors are derived from ColE1 and all are de-

ficient in the *mob* genes and *oriT*, except pBR322, which can be mobilized by RP4. IncQ plasmids such as the 8.6-kb RSF1010 have an extremely broad host range and are mobilized efficiently by plasmids of the IncP group, even into plant cells.

Mobilization of the Bacterial Chromosome

Many conjugative plasmids can mobilize part or even the whole bacterial chromosome with frequencies in the range of 10^{-3} to 10^{-8} per donor cell and this phenomenon has been discovered with the F factor. This plasmid can integrate into the *E. coli* chromosome by one of two different mechanisms: either by homologous or by illegitimate recombination. Homologous recombination is the major mechanism and allows integration at approximately 20 chromosomal sites scattered around the chromosome, making use of the IS elements IS*2* and IS*3* present on both the F plasmid and the *E. coli* chromosome. Illegitimate recombination, the minor mechanism, involves either the IS*2* or the IS*3* element of the F factor and allows integration into the *E. coli* chromosome at any site. *E. coli* strains with an integrated F plasmid are called Hfr (for *h*igh *f*requency of *r*ecombination) and transfer starts at the *oriT* of the integrated F factor. Very few plasmids apart from F can form stable Hfr donors. One exception is the IncJ group of plasmids, which integrate at precise locations in the bacterial chromosome.

The Ti Plasmids

Virulent strains of *A. tumefaciens* are able to transfer DNA from one of their plasmids to plant genomes, which develop tumors called grown galls. This disease results from the transfer and expression of oncogenes from the bacterium to susceptible plant cells. Therefore, the plasmid has been named Ti plasmid (for *t*umor-*i*nducing) and the transferred DNA is T-DNA (for *t*ransfer DNA). As a genus, *Agrobacterium* can transfer DNA to a remarkably broad group of organisms, including numerous dicot and monocot angiosperm species and gymnosperms. Additionally, *Agrobacterium* can transform fungi, including yeast, ascomycetes and basidiomycetes and even human cells. The size of the Ti plasmids varies over 200–800 kb and that of the T-region carrying the T-DNA varies over 10–30 kb. Some Ti plasmids carry one, others multiple T-regions. T-regions are defined by T-DNA border sequences, 25 bp in length, flanking T-regions as direct repeats. The genetic organization of a typical Ti plasmid is presented in Fig. 10.4.

Transfer of the T-DNA occurs upon activation of the *vir* genes. The first step is chemotactic movement of the bacterial cell into a wounded plant, which secretes an exudate. This exudate contains phenolic compounds such as acetosyringone which is sensed by the VirA protein. This protein is an integral inner membrane protein and belongs to the two-component signal-transduction family (see Section 6.2.3). In coordination with the monosaccharide transporter ChvE and in the presence of the appropriate phenolic and sugar molecules, VirA autophosphorylates at an invariant histidine residue and subsequently transphosphorylates the response regulator, the VirG protein. Phosphorylated VirG activates transcription of

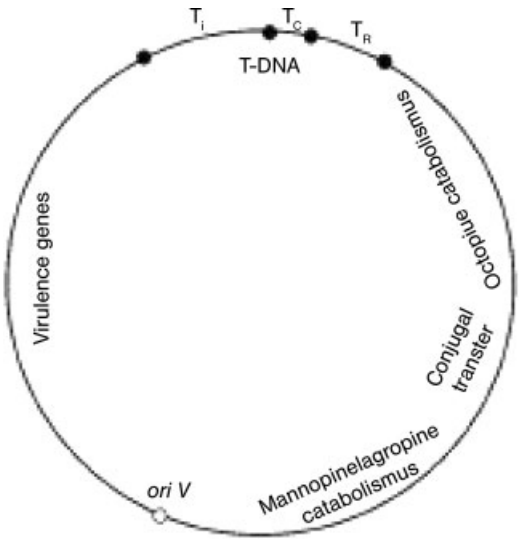


Fig. 10.4 Schematic representation of a Ti plasmid of the octopine type. Shown are the origin of replication (*oriV*), the region of the virulence genes (these are involved in the T-DNA transfer), the T-region, the regions involved in octopine, mannopine and agropine catabolismus and conjugal transfer of the whole plasmid. S.B. Gelvin **2003**, *Microbiol. Mol. Biol. Rev.* 67, 16–37; Fig. 1A.

the remaining *vir* genes by interaction with *vir*-box sequences, being part of *vir* gene promoters. The VirD4 and 11 VirB proteins constitute a type IV secretion system (see Section 8.6.4) involved in the transfer of T-DNA and several other Vir proteins. Most VirB proteins either form the membrane channel or serve as ATPases to provide energy for channel assembly or export processes. Several proteins, including VirB2, VirB5 and possibly VirB7, make up the T-pilus. VirB2, which is processed and cyclized, is the major pilin protein. The T-pilus forms a continuum with the channel formed by the VirB proteins and punctures the plant plasma membrane to deliver the T-DNA and effector proteins into the cytoplasm. The transfer is initiated by the VirD1/VirD2 endonuclease which binds to the right border and introduces a single-strand nick between nucleotides 3 and 4 of the border sequence. Nicking of the border is associated with linkage of the VirD2 protein, through a tyrosine residue, to the 5′ end of the resulting single-stranded T-DNA molecule, termed the T-strand. VirD2 may serve as a pilot protein to guide the T-strand to and through the type IV export apparatus. Since VirD2 contains *n*uclear *l*ocalization *s*ignal (NLS) sequences, it may help direct it and the attached T-strand to the plant nucleus and may play a role in integration of the T-strand into the plant genome. Together with VirD2, VirE2 is transferred into the plant cell. Since it binds unspecifically to single-stranded DNA, VirE2 may keep the T-strand in an elongated shape to help translocation through the nuclear pore and, additionally, may protect the T-DNA from nucleolytic degradation. This view is sustained by the finding that VirE2 also contains NLS sequences.

M. Llosa, F. de la Cruz **2005**, Bacterial conjugation: a potential tool for genomic engineering, *Res. Microbiol.* 156, 1–6.

T. Tzfira, et al. **2004**, *Agrobacterium* T-DNA integration: molecules and models, *Trends Genet.* 20, 375–383.

10.2.2 Gram-positive Bacteria

Conjugation in Gram-positive bacteria has been detected in *E. faecalis*. DNA transfer in other species was reported later followed by evidence for the ability of certain conjugative elements to transfer between different genera. Instead of sex pili, *surface adhesins* are responsible for the initial contact between donor and recipient cells. Furthermore, nonplasmid elements integrated in the chromosome and capable of transferring to recipient cells have now been detected, known as *conjugative transposons* (see Section 10.2.3). Conjugative plasmids generally range in size from about 36 kb to over 100 kb. Some exhibit a broad host range (e.g., pAMβ1), while others have a narrow host range (such as the enterococcal pAD1). Conjugative plasmids are able to mobilize nonconjugative ones, involving cointegration or not, depending on the plasmid. Even mobilization of chromosomal genes is possible. Some plasmids can enhance the transfer frequency of conjugative transposons integrated into the chromosome or inhibit the transfer of a coresident conjugative plasmid.

Conjugative Plasmids in *Enterococcus faecalis*

Conjugative plasmids in *E. faecalis* belong to one of two classes. Those belonging to class I code for *sex pheromones* (including the hemolysin/bacteriocin pAD1 and pOD1, the tetracycline-resistance pCF10 and the bacteriocin pPD1 plasmids), those belonging to class II do not (pAMβ1, pIP800, pFK14). The plasmids of the sex pheromone family in *E. faecalis* are among the most efficient bacterial conjugation systems known. The family consists of over 20 plasmids and shows extensive sequence homologies. Plasmid transfer is highly regulated and only induced by recipient cells producing sex pheromones in close proximity to plasmid donors. Sex pheromones are linear peptides, seven or eight amino acids long, which are produced by plasmid-free recipient cells. In general, *E. faecalis* cells secrete a variety of different sex pheromones simultaneously, such as cAD1, cPD1 and cCF10. The chromosome-inherited pheromone determinants are not linked and code for a lipoprotein precursor with C-terminal seven or eight residues of the related signal sequence corresponding to the mature peptides. The processed lipoproteins have no known function, but resemble the *B. subtilis* SpoIIIJ protein. The sex pheromones bind to donor cells, where they stimulate expression of the *tra* genes encoded by the appropriate plasmid. The pheromone-responding plasmids present in the donor cells encode a surface lipoprotein binding to the cognate pheromone peptide, providing some degree of specificity. This OppA-like protein (for *o*ligo*pep*tide *A*) acts together with an uptake system to transfer the pheromone into the cytoplasm, where it releases a transcriptional block exerted by the PrgX/TraA family of repressors. This leads to an induction of plasmid transfer; and matings performed overnight on solid surfaces can approach 100% transfer. In broth matings, clumping of donor and recipient cells is observed, which is based on the synthesis of a protein called *a*ggregation *s*ubstance (AS) coating the donor

surface and its interaction partner, lipoteichoic acid. The AS encoded by pheromone-responding plasmids such as pAD1 (coding for a hemolysin/bacteriocin), pCF10 (tetracycline-resistance), pOB1 (hemolysin/bacteriocin) and pPD1 (bacteriocin) is a large surface protein with about 1250 amino acids. After uptake of a plasmid, the transconjugant stops producing the sex pheromone either by reducing the synthesis of the pheromone or by producing a plasmid-borne inhibitor peptide, an *antipheromone* that competes with the cognate endogenous pheromone. These hydrophobic inhibitor peptides (e.g., iAD1, iCF10) consist of seven or eight amino acid residues and are processed from 20 to 23 amino acid precursors. Up to three different pheromone-responding plasmids can reside in one single *E. faecalis* cell. When a single pheromone is provided, only the related cognate plasmid is transferred, starting at its *oriT*. Less is known about the genes and functions of the second class of conjugative plasmids of *E. faecalis* that do not respond to sex pheromones.

Conjugative plasmids have also been described but are less studied in the genera *Streptococcus* (pIP501, pIP612, pAC1, pSM15346), *Staphylococcus* (pG01, pMG1, pXU10, pSH8), *Lactococcus* (pNP40, pTN20, pRS01, pIL205), *Listeria* (pWDB100, pIP811), *Bacillus* (pHT73, pXO12, pLS20) and *Clostridium* (pIP401, pCW3, pJU124); and, in most cases, only a few plasmids are mentioned.

Conjugative Plasmids in Streptomycetes

Streptomycetes are spore-forming, mycelial soil bacteria well known as producers of antibiotics. Conjugative plasmids are associated with a phenomenon known as *pock formation*. When a plasmid-containing spore develops within a densely growing population of recipient cells, transfer of the plasmid leads to the formation of growth retardation zones called pocks, where aerial mycelium formation is temporarily inhibited. Plasmid transfer is very efficient, leading to up to 100% of transconjugants. In addition, most plasmids are able to promote the transfer of chromosomal genes; and this mechanism is called *c*hromosome *m*obilizing *a*ctivity (Cma). The mechanism of plasmid transfer is different from that of other plasmids and only one *tra* gene seems to be essential for transfer and Cma activity. In the case of pIJ101, efficient transfer also requires a locus termed *clt* (for *c*is-acting *l*ocus of *t*ransfer), a 54-bp sequence that includes several direct and inverted repeat sequences reminiscent of the *oriT* regions. The amino acid sequence of the Tra proteins contains a conserved nucleotide-binding domain (Walker motifs A and B) and a resemblance to the sporulation protein SpoIIIE of *B. subtilis* and the cell division protein FtsK of *E. coli*. Both of these proteins mediate double-strand DNA transfer; and transfer of pSAM2 and pIJ101 involves double-strand DNA.

The transfer of *Streptomyces* plasmids appears to involve two consecutive steps: (a) *intermycelial* transfer mediated by *tra* and (b) *intramycelial* transfer or movement within the recipient mycelium mediated by *spd* (*sp*rea*d*) genes. Conjugative plasmids include autonomous circular plasmids (pIJ10, pSN22, pJV1) and linear replicons from 11 kb (pSAM2) to 387 kb (pPZG101). Additionally, elements that are integrated in the chromosome and can excise to become autonomous plas-

mids (SLP1, pSAM2; see Section 4.3.5) are widespread. The transfer functions of the linear plasmids are still unknown, but pock formation has been observed.

10.2.3
Conjugative Transposons

While conjugation was originally thought to be exclusively a plasmid-encoded function, it is now known to be mediated by a diverse group of specialized genetic elements, collectively referred to as either conjugative transposons (CTns) or integrative and conjugative elements (see also under Section 4.3.5). These elements can broadly be defined as self-transmissible, integrating elements that cannot be maintained extrachromosomally; and they have now been identified in a wide array of Gram-negative and Gram-positive species from both clinical and environmental sources. CTns were first identified for their role in the dissemination of antibiotic resistance. The first example of a transposon able to move intercellularly was Tn*916*, an element initially identified on the chromosome of *E. faecalis* strain DS16. Closely related transposons have an extremely broad host range of over 35 different genera. There appears to be a strong preference for location on a low copy number replicon, the chromosome or low copy number plasmids. Conjugative transposons are widespread and have been described with *S. pneumoniae* (Tn*5223*; 66 kb; tetracycline and chloramphenicol resistance), *S. pyogenes* (Tn*3701*; >50 kb; tetracycline and erythromycin resistance), *S. agalactiae* (Tn*3951*; 67 kb; tetracycline, chloramphenicol and erythromycin resistance), *L. lactis* (Tn*5276*, Tn5301, Tn5307; ~70 kb; encoding nisin production and able to ferment sucrose) and *C. difficile* (Tn*5397*; 21 kb; tetracycline resistance). *C. difficile* and *C. perfringens* carry transposons that, while not self-transmissible, can facilitate mobilization of a replicon into which they are integrated via the presence of their own *oriT* sites and Mob proteins. It should be mentioned that conjugative transposons have been also found in Gram-negative bacteria, including those of *Bacteroides* species.

A.A. Salyers, N.B. Shoemaker **1997**, Conjugative transposons, *Genet. Eng.* 19, 89–100.

10.3
Transduction

Bacteriophages are viral parasites of bacteria first described independently early in the 20th century by Twort and D'Hérelle. They are among the most abundant groups of replicating entities on our planet. It has been calculated that their number in the ocean probably exceeds 10^{29}. Phages constitute a significant force in bacterial genome evolution and can shuttle large portions of DNA into recipient host genomes. They recognize their host cells via specific receptors located on the cell surface, inject their nucleic acid (usually DNA, sometimes RNA) into the cytoplasm and follow one of two different routes. *Lytic phages* express their genes,

multiply their DNA, package most of the newly synthesized nucleic acid molecules into phage particles and lyse the cells. The T phages of *E. coli* (T1–T7) are well studied lytic phages. A special case of lytic phages are the filamentous phages of *E. coli* such as M13, fd and f1, which do not lyse their host cells. The phage particles are assembled at the surface; and they leave the host cells without lysing them. In contrast, *temperate* or *lysogenic phages* choose between two pathways after infection. They can either enter the *lytic pathway*, leading to multiplication and cell lysis, or alternatively enter the *lysogenic pathway*. The decision to enter the lytic or the lysogenic pathway is made about 10 min after infection, regulated in a complicated way and based mainly on the physiological state of the cell. In short, if the cell is rich in nutrients, the infecting phage will enter the lytic pathway, and if the cell is poor in nutrients (starving), it will enter the lysogenic pathway. If the decision has been made to enter the lysogenic pathway, sufficient amounts of a *repressor* protein will be synthesized, which will bind to one or more operator sequences and thereby repress transcription of those genes involved in lytic development. Afterwards, the phage DNA will be either integrated into the host chromosome (*E. coli* phage λ is the paradigm) or established as a plasmid (e.g., *E. coli* phage P1). Synthesis of the phage repressor continues and the repressed phage DNA is now called a *prophage*. The prophage can enter the lytic cycle either spontaneously or after severe damage of the DNA (exemplified by prophage λ). The first step of this process entails inactivation of the phage repressor, followed by excision from the chromosome. All phages (and viruses in general) carry so-called *early* and *late genes* coding for early and late proteins. After infection, the host RNA polymerase transcribes a few genes called early genes. They code for proteins involved in regulation and in replication. Next, replication of the phage DNA or RNA takes place, followed by expression of the late genes. These code for phage coat proteins (head, tail, tail fibers, etc.) and protein involved in cell lysis. Some phages, such as T4 or T7, code for their own RNA polymerase which transcribes the late phage genes.

Normally, phage DNA molecules are packaged into their heads. But occasionally, they package chromosomal DNA instead of or together with their own DNA. If viable phage particles are formed, they can inject the host DNA into appropriate bacterial cells, where it can recombine with the genome DNA. This process has been called *transduction*. There are two types of transduction: *generalized* and *specialized* transduction. Generalized transduction means that essentially any region of the bacterial DNA can be transferred from one cell to another; and this process can be carried out by both lytic and lysogenic phages. In contrast, specialized transduction is performed by lysogenic phages only. Here, excision of the prophage DNA does not occur precisely; and neighboring chromosomal DNA is excised, together with prophage DNA. This hybrid DNA is then replicated and packed into phage particles. Two types of phage particles can result: *viable, infectious* and *nonviable, defective*. Viable particles carry all the phage genes responsible for successful propagation. In contrast, defective particles lack at least one essential gene necessary for successful propagation due to the imprecise excision event; and they need coinfection with a helper phage.

10.3.1 Generalized Transduction

In generalized transduction, virtually any bacterial gene can be transferred from the donor to the recipient cell. Generalized transduction was first detected in *S. typhimurium* with phage P22 and later with phage P1 in *E. coli*. How do generalized transducing phage particles arise? During lytic infection, the enzyme responsible for packaging viral DNA molecules into phage heads accidentally packages host DNA molecules of about the same size as the phage DNA. In the case of P1, about 100 kb are packaged, in the case of P22 44 kb. Therefore, any viral lysate of generalized transducing phages contains, besides normal phage particles, a few transducing viruses (in the range of 10^{-6} to 10^{-8}). Upon infection of a sensitive host cell, the transducing DNA is released and can undergo genetic recombination with the host chromosome if it is a piece of the bacterial chromosome. If it is a plasmid, it may circularize, using direct repeats of sufficient length. If the incoming DNA contains a transposon, the mobile element may transpose into the chromosome or a plasmid if present. Cells that have received DNA from another bacterium are called *transductants*.

Why is one phage a generalized transducing phage, while another is not able to pick up and transfer DNA? Packaging of many phage DNAs relies on a so-called *pac* site which is recognized by a packaging enzyme. It will carry out a double-strand cleavage at the *pac* site and, using the linear end, it starts packaging of the viral DNA into proheads. Since most phages follow the headful packaging mechanism, the prohead will be filled with DNA before the enzyme carries out the second double-strand cleavage. Then, the new linear end of the concatemer will be used to start packaging into another empty prohead, etc. In the case of P22, it has shown that, from a single *pac* site, about ten headfuls of DNA can be packaged before the mechanism requires another *pac* site. If the bacterial chromosome or plasmids present carry *pac*-like sites, these will be recognized by the enzyme, resulting in the packaging of bacterial DNA into proheads. With both P1 and P22, the *pac* sites are not very specific, allowing the formation of generalized transducing particles. Virulent phage T4 does not package host DNA, since it will degrade it completely. If this is prevented by appropriate mutations, T4 acts as an excellent generalized transducing phage. Since its DNA does not contain *pac* sites, it packages any DNA with high efficiency.

A special group of transducing phages has been described where all phage particles contain host DNA. These phages reside as prophages in the bacterial chromosome and are induced by treatment with mitomycin C in the case of VSH-1 present in the anaerobic spirochete *B. hyodysenteriae*. VHS-1 prophage genes span a 16.3-kb region and are organized into head (seven genes), tail (seven genes) and lysis modules (four genes) representing only about 50–60% of the genome of self-replicating prophages. Purified VSH-1 virions are noninfectious, contain random 7.5-kb fragments of the bacterial genome and no phage DNA and mediate generalized transduction of *B. hyodysenteriae* cells. Another example is the phage GTA of *R. capsulatus*. The prophage codes for 19 ORFs (15 kb) for putative head and tail

functions. The virions carry random 4.5-kb fragments of host genome and are noninfectious, with generalized transduction ability.

10.3.2 Specialized Transduction

Specialized transduction is defined as transduction of specific genes at a high frequency. It is dependent on temperate phages which integrate their DNA into the bacterial chromosome at defined sites. Occasionally, the prophage DNA is excised incorrectly, carrying chromosomal DNA from one or the other end of the prophage. These DNA molecules are replicated and packaged into phage heads leading to phage particles. Upon infection of host cells, the injected phage DNA replicates in the case of viable specialized transducing phages and in the presence of helper DNA in the case of nonviable phages with defective replication functions. This transducing DNA can either integrate again as prophage together with the attached bacterial DNA or, alternatively, homologous recombination can occur between the bacterial DNA of the phage DNA and homologous regions on the bacterial chromosome.

The best studied specialized transducing phage is the temperate phage λ of *E. coli*. λ DNA integrates into the *E. coli* genome by site-specific recombination between *attP* and *attB* sites. While *attP* is located on the λ DNA, *attB* is present on the *E. coli* DNA sandwiched by the *gal* and *bio* operons. Both sites contain an identical core DNA sequence of 15 bp (Table 10.1). Besides *attB*, which serves as the major integration site (99% of the λ DNA molecules recombine with *attB*), several minor integration sites have been identified (Table 10.1). These sites are distinguished from *attB* by 2–8 deviations within the core sequence.

When lysogenic bacteria are induced by UV light or certain chemicals, such as mitomycin C, the prophage DNA is usually excised at both ends of the prophage by site-specific recombination between *attL* and *attR* sites. These integrative and excisive recombination events are mediated by the phage-encoded proteins Int (integrase), Xis (excisionase) and the two host-encoded accessory proteins IHF and Fis. While Int and IHF are necessary for both reactions, Int, Xis and Fis promote excision. These proteins form higher-order complexes with the *att* sequences for the promotion of integration and excision. Occasionally, *E. coli* DNA adjacent to

Table 10.1 Known integration sites of phage λ.

Site	*Core DNA Sequence*		
attP	TCA*GCTTT*	*TTTATAC*	*TAAGTTGG*
attB	CCT*GCTTT*	*TTTATAC*	*TAACTTGA*
proB	TGC*GCTAA*	*TTTATAC*	GAGGC*TAC*
trpC	GCGTAA*TG*	*TTTATAA*	ATGGCG*GC*
galT	CGCCT*TTG*	*TTTTCAA*	AAACC*TGC*
thrA	CGG*GCTTT*	*TTTCTGT*	GTTTCCTG
rrnB	TTG*GCTAT*	*TTTACCA*	CG*ACTGTC*

the prophage DNA becomes recombined with phage DNA, and transducing phages that contain the bacterial genes *gal* or *bio* (replacing some of the phage DNA) are produced. If they contain genes from the *gal* operon, they are called λ*gal*; if they carry genes from the other operon, λ*bio*. Nucleotide sequence analysis of the junctions of λ*bio* transducing phages indicates that the recombination can take place at many sites on both the bacterial and phage genomes and that the recombination sites contain short regions of homology, of 5–14 bp. Therefore, the abnormal excision which occurs during the formation of specialized transducing phages is caused by illegitimate recombination, which can be induced by UV irradiation or other DNA-damaging agents and is independent of RecA.

C. Canchaya, et al. **2003**, Phage as agent of lateral gene transfer, *Curr. Opin. Microbiol.* 6, 417–424.

10.4
Horizontal Gene Transfer

*H*orizontal *g*ene *t*ransfer (HGT), also called lateral transfer, is the intra- and interspecies exchange of genetic information and plays an essential role in the evolution of bacteria. HGT is also known to occur in eukaryotes. For example, DNA transposons have been found to be horizontally transferred between species of the fruitfly *Drosophila*. The first evidence that HGT could occur was the observation that virulence determinants could be transferred between pneumococci in infected mice by transformation. Another important observation was the spread of antibiotic resistance markers within and among species, starting in the early 1950s. Now we know that the diversification of bacterial species is largely caused by horizontal transfer of diverse mobile DNA elements, such as insertion elements, transposons, integrons, genomic islands (groups of chromosomal genes transferred together), plasmids and combinations of these elements. Compositional analysis of bacterial chromosomes has revealed that considerable proportions of most bacterial genomes consist of horizontally acquired genes. For example, *E. coli* and *S. typhimurium* are separated about 100×10^6 years ago. While ~90% of the housekeeping genes are identical, about 17% of their genomes (~800 kb) have been acquired by HGT.

The three major mechanisms (transformation, transduction, conjugation) provide bacterial populations with access to a horizontal gene pool, enabling them to quickly respond to environmental challenges. HGT is considered a ''quantum leap'' in evolution, as the genes are generally transferred *en bloc*, forming islands of heterologous DNA in the new host. One example is clusters of virulence genes present in many bacterial pathogens but not in closely related nonpathogenic strains or species. These gene clusters may be located on transmissible phages or plasmids, but are often found as so-called *pa*thogenicity *i*slands (PAIs) on the chromosome. However, specific transmissibility of PAIs has not yet been demonstrated and their evolutionary origin remains unknown. The PAIs display the fol-

lowing characteristics: (a) chromosomal segments of 35–45 kb, (b) carry virulence genes, (c) G+C content differs from the remainder of the chromosome, (d) inserted at the 3′ end of a tRNA gene, (e) flanked by repeated sequences.

Horizontally acquired genes can cause one of three effects:

- They cause deleterious effects in the chromosome of the bacterial recipient; and these bacteria are lost from the population over time.
- Some horizontal acquisitions might be effectively neutral; and their survival depends on chance events.
- Horizontally transferred genes that confer a selective advantage to the host and mobile elements that encode their own transfer and maintenance functions have the potential to spread rapidly within a bacterial population.

Based on the transformation mechanism of gene transfer, we would like to suggest a new possibility to increase the bacterial diversity:

- Some cells in a starving bacterial population become hypermutable and acquire numerous point mutations per chromosome (see Section 5.2.4).
- These cells either secrete their DNA or induce self-lysis (this has been described to occur).
- The free DNA is taken up by other cells of the population of the same or closely related species which can acquire favorable mutations.

Stability of Free DNA in the Environment

The persistence of free DNA in the environment plays an important role in transformation. The longer the DNA survives, the higher the chance for being taken up. In soil and sediments, about 1 μg of DNA can be recovered from 1 g material, while in fresh and marine water approximately 0.03–88.0 μg DNA was measured per liter. The addition of pure DNA into soil, water or various types of food and feed revealed that the DNA is not immediately degraded and will persist for hours to days.

Barriers to HGT

A number of processes limit the transfer, uptake and stabilization of foreign DNA molecules in the same or in foreign bacterial species. Whereas in transformation and transduction the recipient exerts the more active role, in conjugation the donor seems to play the positive role. *E. coli* cells carrying the F factor are poor recipients for another copy of the F plasmid. This phenomenon has been called *surface exclusion* and seems to create an effective barrier against conjugative transfer into cells carrying already a copy of the plasmid (see above).

Another barrier against foreign DNA are restriction enzymes. DNA that is recognized as foreign because of its lack of appropriate modification can be cleaved into fragments by restriction endonucleases. For small plasmids and individual

genes, there is a significant chance that they might be free of recognition sequences. Furthermore, DNA that enters recipient cells via conjugation or transformation is normally single-stranded and will provide some protection. Broad-host range IncP-1 plasmids such as RP4 have adapted to the existence of such barriers by the selection of variants devoid of most sites. Another strategy uses genes encoding products that interfere with type I restriction systems. Orthologs of the *ardA* gene are found quite commonly in conjugative bacteria.

Host Range

Early investigators recognized that host range was an intrinsic property of a bacterial plasmid. Their studies established that some replicons could only be maintained in a single or closely related bacterial host and thus had a narrow host range, while other plasmids were promiscuous and could replicate in unrelated bacterial genera. While the host range of a plasmid can be influenced by factors affecting its capacity to be transferred into a new host, after entry it is primarily the plasmid-encoded replication system and its interaction with host factors which determines the ability of a plasmid to maintain itself in the recipient cells.

H. Philippe, C.J. Douady **2003**, Horizontal gene transfer and phylogenetics, *Curr. Opin. Microbiol.* 6, 498–505.

C.M. Thomas, K.M. Nielsen **2005**, Mechanisms of, and barriers to, horizontal gene transfer between bacteria, *Nat. Rev. Microbiol.* 3, 711–721.

J.P. Gogarten, J.P. Townsend **2005**, Horizontal gene transfer, genome innovation and evolution, *Nat. Rev. Microbiol.* 3, 679–687.

Glossary

AAA$^+$ (for *A*TPases *a*ssociated with a variety of cellular *a*ctivities). These ATPases function as mechanochemical machines to remodel their protein substrates. They contain a highly conserved AAA$^+$ module of about 230 amino acids that is present in one or more copies in each protein. Each module can typically be divided into two subdomains: one ATPase and one Zn^+-binding domain. In the ATPase subdomain, there is a conserved Walker A motif, involved in binding the phosphate of ATP, and a Walker B motif, involved in metal binding and ATP catalysis. The unifying structural feature of these AAA$^+$ proteins is the arrangement of the subunits into ring-shaped hexameric or heptameric complexes. Proteins with the AAA$^+$ motif use conformational changes derived from ATP hydrolysis to remodel client proteins.

Adaptive response. Activation of the genes of the Ada regulon through methylated Ada protein.

Allele. One of the numerous forms of a gene. We distinguish the wild type from mutant alleles.

Allele-specific suppressor. A second-site mutation in another gene that restores the wild-type phenotype.

Amphipathic peptides. Peptides containing both hydrophilic and hydrophobic amino acid residues. The spatial separation of these residues facilitates their attachment and insertion into membranes.

Anti-anti-sigma factor. A protein that counteracts the action of its cognate anti-sigma factor.

Antisense RNA. A naturally occurring short untranslated RNA transcript that can base-pair with a complementary mRNA to prevent its translation.

Anti-sigma factor. A negative transcriptional regulator acting by binding to its cognate sigma factor to prevent its interaction with the core RNA polymerase.

Antitermination. The default pathway results in premature termination; and a regulatory molecule promotes transcription read-through.

ATP-binding cassette (ABC) transporters. These transporters couple the energy released from the hydrolysis of ATP to the translocation of a wide variety of substances in or out of cells.

Attenuation. In transcription attenuation, the default pathway is read-through and a regulatory molecule induces transcription termination.

Dynamics of the Bacterial Chromosome. Structure and Function. W. Schumann
Copyright © 2006 Wiley-VCH Verlag GmbH & Co. KGaA
ISBN 3-527-30496-7

Autocleavage. A process by which a protein cuts itself such as the LexA repressor after extensive DNA damage.

Autoinducer. A small molecule that allows intercellular chemical communication by bacteria. Autoinducers play an important role in quorum sensing. Gram-negative bacteria produce acylated homoserine lactone (AHL) autoinducers which are composed of an acyl chain attached to a homoserine lactone. Gram-positive bacteria produce peptides as autoinducers which are either taken up into the cell through an oligopeptide uptake system or bind to a sensor kinase.

Autophosphorylation. A process by which a protein (e.g., sensor kinase) transfers a PO_4 group from ATP or another phosphate source such as acetyl phosphate to itself.

Branch migration. The process by which the site at which two double-strand DNA molecules held together by crossed-over strands (Holliday junction) moves in one direction or the other.

Chaperones. A class of proteins that bind to partially or completely denatured (non-native) proteins, thereby promoting their proper folding during synthesis or after damage (folder chaperones) or preventing them from forming aggregates (holder chaperones).

Chi. An octanucleotide (in *E. coli*) that stimulates DNA repair by homologous recombination.

Closed complex. The closed complex is formed between RNA polymerase and duplex promoter DNA preceding the initiation of DNA untwisting.

Coiled-coil motif. The α-helical coiled-coil motif contains heptad amino acid repeats (*a-b-c-d-e-f-g*), where a and d residues are hydrophobes, providing an apolar stripe. The heptad repeat-containing subunit α helices wind around each other and pack their side chains in a knobs-into-holes manner, providing the left-handed superhelical coiled-coil structure.

Cointegrate. Covalent fusion of two different replicons such as fusion of two different plasmids.

Competence. The physiological state of prokaryotes enabling them to take up DNA fragments from the environment and sometimes to stably incorporate them into their genomes.

Condensin. Protein that binds chromosomal DNA in two different places, folding it into large loops and thereby condensing it. Examples are MukB in *E. coli* and Smc in *B. subtilis*.

Conjugation. Transfer of genes is mediated by certain plasmids or integrative conjugative elements (ICEs) with relevant transfer genes. Cell–cell contact is required for conjugation.

Conjugative transposons (*CTns*). Transposons that code for functions allowing then to transfer themselves into other bacterial cells.

Cooperative binding. The binding of one protein molecule to a DNA or RNA site greatly enhances the binding of one or more additional protein molecules to adjacent sites.

Decatenation. The process of separating two DNA daughter molecules by topoisomerases.

Dimerization domain. The region of a protein that allows interaction with another molecule of the same (homodimer) or a different (heterodimer) protein molecule.

Disulfide oxidoreductase. An enzyme that can catalyze either formation or breakage of disulfide bonds between cysteine residues.

D-loop. A recombination intermediate formed by base-pairing of a ssDNA with complementary bases in a dsDNA molecule, thereby displacing one strand (displacement loop).

Epistasis. A type of interaction where a mutation in one gene can affect the phenotype of another gene.

Essential genes. Genes that are required for survival and growth under given conditions. While a knockout in a gene involved in the biosynthesis of an amino acid is lethal in minimal medium, it is dispensable in rich medium.

GAF domain. Ubiquitous signaling modules found in all three kingdoms of life that bind small molecules such as cyclic monophosphates and formate. They are found in some bacterial enhancer binding proteins and in sensor kinases as part of two-component signal transduction systems.

Genomic islands. Chromosomal regions that were horizontally acquired but that are no longer or were never self-transmissible.

Helicase. A protein that uses the energy of ATP hydrolysis to unwind dsDNA, therefore separating the strands.

Hemimethylated DNA. Only one strand is methylated, e.g., immediately after replication.

Holliday junction. An intermediate in homologous recombination in which one strand from each of two DNAs crosses over and is ligated to the corresponding strand in the opposite DNA.

Host range. All bacterial species and strains in which a plasmid and phage can replicate.

Housekeeping functions. These are functions which are needed mainly during exponential growth, such as transcription, translation, replication, recombination and cell envelope synthesis.

Illegitimate recombination. Recombination between DNA sequences that share little or no homology.

Incompatibility group. All plasmids are grouped into incompatibility groups; and plasmids belonging to the same incompatibility group are not able to coexist without selective pressure within the same cell.

In-frame deletion. A deletion mutation in a gene coding for a protein that removes 3 bp or a multiple of three. Such deletions are not polar.

Insertion sequence. A small DNA sequence (700–2000 bp long) of mobile DNA flanked by inverted repeats encoding normally just one protein, the transposase, catalyzing transposition of the element.

Integrative conjugative elements(*ICEs*). Chromosomally located gene clusters including conjugative transposons (CTns) and genomic islands that encode phage-linked integrases and conjugation proteins together with genes associated with an observable phenotype, such as symbiosis and virulence. ICEs and CTns

can be transferred between cells, whereas genomic islands have not been shown to transfer.

Integron. A genetic element that encodes an integrase enzyme able to assemble tandem arrays of genes and provide them with a promoter for expression. Integrons are often associated with antibiotic multi-resistance.

Intragenic suppressor. A suppressor mutation that occurs in a gene already carrying a mutation and thereby restores a wild-type phenotype.

Isomerization. Step during initiation of transcription leading to strand opening within the –10 region by the RNA polymerase.

Knockout mutation. A mutation (normally a deletion or an insertion) completely eliminating the function of the gene.

Lipoteichoic acids. Teichoic acids with lipid modifications that allow association with the cytoplasmic membrane.

Methly-accepting chemotaxis protein (*MCP*). A transmembrane chemoreceptor that shows methylation-dependent adaptation.

Methyl-directed mismatch repair system. A repair system in enteric bacteria recognizing mismatches in newly replicated DNA and specifically removing the wrongly incorporated nucleotide form the daughter strand. This strand is recognized due to the absence of methylation at nearby GATC sequences.

Min system. A group of two or three proteins that inhibit unwanted formation of Z rings at the cell poles. In their absence, minicells are produced.

Mobile group II intron. A catalytic RNA molecule that acts as a mobile genetic element. It codes for a reverse transcriptase and can insert site-specifically into target DNA.

Mobilization. A plasmid unable to self-transfer can use the *trans*-acting proteins of a conjugative plasmid residing within the same cell to be mobilized into a recipient cell.

Molten globule. State of a protein with secondary but devoid of tertiary structure.

Moonlighting proteins. These are bifunctional proteins.

Non-homologous joining. Pathway that is used to repair chromosomal double-strand DNA breaks. The process is non-homologous because adjacent strands are fused by direct end-to-end contact without sequence homology. Therefore, non-homologous end-joining is error-prone as it results in a joining of breaks without a template.

Nucleoid. In bacteria, the nucleoid is defined as containing the chromosome and its associated proteins.

Okazaki fragments. Extension products of the RNA primers complementary to the lagging strand, 1000–2000 bp in length.

Open complex. An open complex is formed when the RNA polymerase binds to a promoter and the duplex around the transcription start site is unwound.

Orthologues. Homologous genes with related function in several species, e.g., *recA* and *dnaA*.

Paralogues. Genes that have originated from a common ancestral gene by a duplication event. For instance, *E. coli* codes for seven *csp* (*c*old *s*hock *p*roteins) paralogues.

PAS domain. PAS stands for PER, ARNT and SIM and denotes a protein domain that is involved in recognizing stimuli such as light, oxygen, redox potential, the energy status and small ligands.

Pathogenicity island. A continuous block of genes, of which at least one subset codes for virulence factors.

PDZ domain. PDZ domains are ~100-amino-acid modules initially found in a number of eukaryotic cell junction-associated proteins, where they serve to localize ion channel proteins within a particular region of the membrane by binding to the C termini of the channel proteins. PDZ domains occur in many proteins and mediate specific protein–protein interactions and bind preferentially to the C-terminal three or four residues of the target protein. Crystal structural studies of several PDZ domains have revealed that the C termini of their partner proteins bearing specific sequences bind to a surface cleft in the PDZ domain, with the C-terminal carboxylate group bound by the highly conserved GLGF motif of the PDZ domain through hydrogen bonds to the backbone amides.

Porin. A water-filled protein channel across the outer membrane of Gram-negative bacteria that allows the diffusion of hydrophilic molecules up to a molecular mass of about 600 Da into the periplasm.

Photolyase. An enzyme that uses energy of visible light to split pyrimidine butane dimers in DNA.

Primer. A ssDNA or RNA that can hybridize to a single-stranded template DNA and provide a free 3' hydroxyl end used by DNA polymerases to add dNTPs to synthesize a DNA molecule complementary to the template DNA.

Primosome. A complex of proteins involved in the synthesis of primers for the initiation of DNA strands.

Promoter occlusion. RNA polymerase initiating transcription at an upstream promoter blocks the access of another molecule to downstream promoter(s).

Pseudogene. A DNA sequence that was originally derived from a functional protein-coding gene and has lost its function, owing to the presence of one, or more, inactivating mutations.

Quorum sensing. The ability of bacteria to sense their own cell density by measuring the concentration of signaling molecules that have been released by bacteria into their environment.

Regulon. At least two operons that are regulated by the same transcriptional regulator, an alternative sigma factor, a transcriptional repressor or activator.

Relaxosome. A protein complex bound to the *oriT* of a plasmid and involved in the sequence-specific cleavage of one strand.

Replisome. A complex of DNA replication enzymes consisting of two DNA polymerases for leading- and lagging-strand synthesis, sliding clamps to maintain processivity, a sliding-clamp loader, a helicase to unwind the parental dsDNA and a primase to synthesize RNA primers.

Resolvase. A site-specific recombinase that breaks and rejoins DNA at *res* sequences in the two copies of the transposon in a cointegrate into separate DNAs.

Response regulator. Part of the two-component signal transduction system that controls gene expression in response to external signals. Most response regulators

consist of two domains, a DNA-binding domain and a regulatory domain, the activity of which is modulated by the external signal through phosphorylation by the sensor kinase.

Retrotransposon. A type of transposon that transposes to a new location in a DNA molecule by first making an RNA copy of itself, followed by making a cDNA using reverse transcriptase and then inserting the dsDNA copy into the target DNA.

Reverse gyrase. An enzyme that is responsible for generating positive DNA supercoils in the chromosome of thermophilic Archaea and Eubacteria.

Riboswitches. Highly structured domains within mRNAs that precisely sense metabolites and control gene expression. These RNA elements located in the 5' untranslated regions of transcripts can adopt different mutually exclusive structures. One of these structures is thermodynamically favored, while the alternative fold depends on the binding of an effector and metabolites such as vitamins or amino acids. Binding of the metabolite leads to transcription termination preventing expression of the downstream genes.

RNA chaperones. Ubiquitous RNA-binding proteins promoting activities such as RNA–RNA annealing and strand transfer or exchange.

Sacculus. A synonym for the `sac-like' peptidoglycan molecule surrounding the cytoplasmic membrane of a bacterium.

Sensor kinase. Part of the two-component signal transduction system which senses and transmits external signals to the response regulator.

Signal recognition particle (*SRP*). An RNA/protein complex involved in inserting integral cytoplasmic membrane proteins into their compartment.

SOS response. Inducible repair system in bacteria invoked in response to a sudden increase in DNA damage.

Stringent response. If bacteria are starved for a required amino acid, macromolecular synthesis ceases. The stringent response is signaled by the production of ppGpp by the RelA protein when an uncharged ribosome binds to the A-site of a ribosome.

Supercoiling. The DNA is twisted about itself in the opposite direction (negative supercoiling) or in the same direction (positive supercoiling). Almost all DNA molecules are negatively supercoiled. Supercoiled DNA molecules have a higher energy than relaxed DNA and this energy can alter local DNA structure.

Teichoic acids. Phosphate-rich, anionic polysaccharides that are attached to the peptidoglycan of Gram-positive bacteria.

Transconjugants. Bacterial recipient cells that have received a plasmid or part of a chromosome by the process of conjugation.

Transcriptional antitermination. The default pathway results in premature termination and a regulatory molecule promotes transcription read-through.

Transcriptional attenuation. The default pathway is read-through and a regulatory molecule induces transcription termination.

Transduction. Gene transfer mediated by certain phages. Generalized transducing phages can take any gene of the bacterial chromosome with about the same

frequency, while specialized transducing phages incorporate only a few genes with a high frequency into the same genome.

Transformation. Uptake of naked DNA from the environment.

Transglycosylase. An enzyme that catalyzes the attachment of a peptidoglycan disaccharide-pentapeptide precursor molecule to an existing glycan strand by a β-1,4 glycosidic bond.

Translational coupling. A gene arrangment in which the translation of the first gene is required to allow translation of the second, downstream gene.

Translesion synthesis (*TLS*). Synthesis of DNA over a template region containing a damaged base unable of proper base-pairing.

Transpeptidase. An enzyme that catalyzes the formation of a peptide bond between adjacent polypeptide side-chains, forming a flexible peptide bridge between the glycan strands.

Trigger factor. A chaperone of the holder type bound close to the exit tunnel of the ribosome and interacting with emerging nascent polypeptide chains to prevent their premature folding.

Tumor-inducing (Ti) plasmid. A group of plasmids responsible for transferring DNA, the T-DNA, from *Agrobacterium* to plant cells.

Two-component signal transduction system. Consists of two interacting proteins, a sensor kinase and a response regulator, forming the central signaling pathway in bacteria.

UP element. AT-rich DNA element found upstream of some promoters that enhances promoter activity by providing a binding site for the C-terminal domain of the α-subunit of the RNA polymerase.

Walker A and B motifs. Conserved amino acid sequence signatures that are characteristic of nucleotide binding folds. Occur in all proteins of the AAA and AAA^+ families.

Subject Index

a

ABC, *see* ATP-binding cassette 6
ABC transporter 7
acclimation phase 349
acid stress 367
acid-resistance system 367
Actinetobacter 382
Ada protein 141
adaptive amplification 131
adaptive mutation 129
adaptive point mutation 130
adaptive response 142
adaptor protein 298, 302, 304
Agrobacterium 26
alkaline stress 369
Ames test 128
Anabaena 47
anti-activator 199
anti-anti-sigma factor 178
antimutator genes 124
antiport 6
antiporter 366
anti-sigma factor 177
anti-termination 180, 185
anti-terminator 176
antitoxin 79, 224
aptamers 229
aptazyme 227
aquaporin 354, 360
AR, *see* acid-resistance system 367
Arabidopsis 76
arrest 173
ATP hydrolysis 277
ATP synthase 5
ATP-binding cassette 6
ATP-dependent protease 271, 298
ATP-independent protease 305
attenuation 180
autoinducer 214
autoinducing peptide 217
autoregulatory protein 221
Azotobacter 199

b

bacterial intercellular communication: quorum sensing 214
Bartonella 369
base excision repair 136, 137
BER, *see* base excision repair 136
bgl operon 159
Borrelia 29
Brachyspira 400
Bradyrhizobium 264
Brucella 369
Buchnera 31
bypassing 259

c

Campylobacter 267
Candida 207
CAP 195
capsular polysaccharides 20
capsules 20
Caulobacter 369
cell cycle 51
cell division 51, 73
– assembly of cell division proteins 78
– divisome 73
– postsegregational killing 80
– proteins 78
cell lysis 399
cel operon 160
centromere 71
chaperones 271, 315
– disaggregating 287
– folder 272
– holder 279
– intramolecular 289
chaperonin 191, 275
– functional domains 276
chemotaxis 167, 178, 207–209, 211
Chi 85
Chlamydia 267
chromosomal DNA 31
chromosomal rearrangement 44
chromosomal replication 65
chromosome
– antitoxin 79

Dynamics of the Bacterial Chromosome. Structure and Function. W. Schumann
Copyright © 2006 Wiley-VCH Verlag GmbH & Co. KGaA
ISBN 3-527-30496-7

– circular 29, 51
– condensins 71
– deletion 46
– genes 36
– insertion elements 41
– linear chromosomes 29, 67
– repetitive sequences 36, 37, 40
– replication 51
– retroelements 44
– separation of daughter chromosomes 64
– size 31
– structure of 29
circular protein 268
clamp 57
Clostridium 59
cointegrate 98, 100, 102, 116
cold acclimation 352
cold shock 341, 349
– acclimation phase 349
cold shock degradosom 352
cold shock genes 352
– regulation 352
cold shock proteins 349
– function 352
coldspot 124
competence 382
condensin 72
conjugation 381, 388
conjugative plasmid 396
conjugative transposons 118, 398
conservative integration 115
contrahelicase 63
corepressor 191
crescentin 2
CRP, *see* CAP
cryptic cleavage sites 298
cryptic gene 159
curli fibers 25, 330
curli proteins 26, 320
cytoskeleton 2

d

Dam methylase 157, 260
daughter chromosome 68
Dcm methylase 134, 157
deaminase 367
decarboxylase 366
degradosome 234
Deinococcus 151, 382
deletion 45
desaturase 349
dif 101
disulfide bond 293
divergon 188, 192, 330, 344
divisome 73
DNA
– bent segment 154
– conjugative 388
– damage: tolerance mechanism 150
– damaging agents 131
– deletions 123
– elongations 123
– gyrase 56
– insertions 123
– interstrand crosslinks 150
– inversions 123
– lesions 149
– ligase 57
– looping 155
– persistance 383
– polymerase 56, 59
– repair mechanisms 131
– repair 134
– silencer 159
– supercoiling 31, 154
– template DNA 62
– uptake 383
DNA bending: junction model 154
– wedge model 154
DNA repair: base excision repair 136
– methyl-mediated mismatch repair system 132
– nucleotide excision repair system 134
– repair of oxidized nucleotides 138
– short patch repair 133
DNA strand 56
– lagging strand 56
– leading strand 56
DNA-binding proteins 32
DNA-dependent RNA polymerase 162
double-strand break 62
drop-off 249
Drosophila 101
dual regulators 188
duplication 44

e

ebg operon 161
ECF sigma factor 167
effector 331
– proteins 330
electron transport chains 366
endoribonuclease 233
enhancer 203
Enterobacter 105
Enterococcus 366
enterotoxin 337
envelope 376
environment 382
epigenetic switch 158
error prone 95
Erwinia 105
Euprymna 215
extended –10 region 165, 177, 185
extragenic 122

f

fast-stop mutant 54
Fenton reaction 138
fimbriae 106
flagella 1, 21, 178
– composition of 21, 22
flagellin 104, 332

flagellum 331
– flagellar filament 331
– flagellin subunits 332
– hook 331
frameshift mutations 121, 122

g
gated channels 17
gene amplification 45
– cassettes 107, 110
– cryptic 159
– mutator genes 124
– overlap 263
– regulation 153
genes 36
– general stress 341
– organization 36
– pH regulation 364
– pseudogenes 37
genes within genes 259
genetic competence 381
genome 29
– size 31
global regulator 220
group II intron 44
GreA, GreB 173, 185
GroEL chaperonin 163, 193,
GTP hydrolysis 320

h
Halobacterium 267
Haloferax 76
heat sensors 343
heat shock 282, 341
– gene 228
– protein 281, 282
heat shock proteins 342
– function 343
helical pitch 154
Helicobacter 382
HGT, *see* horizontal gene transfer 29
Holliday junction 85, 90–92, 149
homeoviscous adaptation 349
homing 110, 112
– endonuclease 110
Hoogsten base pair 139
horizontal gene transfer 29, 402
host defense system 306
hotspot 124
hydroxyl radicals 361
hyperosmotic shock 354
– adaptation 355
– response 354
hypoosmotic shock 354
– response 359

i
IHF, *see* integration host factor 33
illegitimate recombination 112
incompatibility 389
injectisome 286, 333
inner membrane 323
– insertion proteins 323
insertion sequence 113
integrase 98, 106
integrating conjugative elements 118
integration host factor 33
integrons 107, 108
– antibiotic resistance 109
– gene cassettes 107
intragenic suppressor 122
intrinsic terminators 175
intron 110, 239
inversion 49, 107
inversion systems 103, 104
– shufflons 107
invertase 98
invertasome 104
IS elements 113
IS, *see* insertion sequence 113
isomerization 170
iterons 54

k
kissing complex 225
Klebsiella 44
Kleisin 72

l
λ 99
lac operon 191
LacI 129
LacI test 128, 129
Lactobacillus 240
Leishmania 154
leucine zipper 76
lipid II 9,
lipopolysaccharide 10, 14, 16
lipoprotein 12
Listeria 135
lytic phage 399

m
magic spot 171
maltose 7
maltose uptake system 8
mating aggregates 391
mating junctions 391
MDOs 15, 16
mechanosensitive channel 359
Mesorhizobium 369
metabolic stress 371
Methanococcus 76
methylase 59
methylation 59
methyl-mediated mismatch repair 132
Mfd factor 174
mimicry 221
MMR, *see* methyl-mediated mismatch repair 132
molecular chaperone 271
molecular crowding 290
moonlighting proteins 259, 266
mRNA degradation 237

– decay 230
mucosal invasion 225
multiresistance integron 109
mutagen 121, 126
– chemicals 126
– high-energy radiation 126
mutagenesis
– mutagenic step 124
– premutagenic step 124
mutagenic nucleotide 125
mutasome 95
mutation 121
– adaptive gene amplification 130
– adaptive point mutation 130
– adaptive response 141
– avoidance mechanisms 131
– conservative 122
– detection systems 128
– experimentally enhanced 126
– frameshift 122
– induced 121
– mutagenic agents 126
– nonsense 122
– origin 124
– point 121
– rearrangements 123
– silent 122
– SOS response 143
– spontaneous 121
– UV irradiation damage 140
mutator 124
mutator genes 124
Mycobacterium 135

n
natural selection 121
ncRNA 220, 222, 243, 255, 362
N-end rule 298
NER, *see* nucleotide excision repair 134
Neurospora 207
NHEJ, *see* nonhomologous endjoining 119
nonhomologous endjoining 119
nonstop mRNA 251
nucleoid 31, 33
nucleotide excision repair 134, 135

o
Okazaki fragment 52, 58
open reading frame 48
ORF, *see* open reading frames 36, 40
osmolality 353
– hyperosmotic shock 354
– hyperosmotic 354
– hypoosmotic shock 354
– hypoosmotic 354
osmoprotectant 356
osmoregulation 209
osmotic pressure 355
osmotic shock 353
outer membrane 9, 324
overlapping genes 259
oxidative shock: response 361
oxidative stress 227, 341
oxygen radical attack 361

p
Par system 72
partitioning 68, 73
– postseptational 73
partner-switching module 178
pathogenicity island 332, 402
peptide pheromone 214, 217
peptidyl tRNA hydrolyse 249
periplasmic proteins 15, 16
periplasmic space 15
pH homeostasis 365
phage 99
phosphorelay 207
pili 23, 284, 390
– biogenesis 23
– curli 25
– T-pilin subunits 27
– T-pilus 26
plant pathogenesis 225
plasmid 29, 54, 68
– conjugative 394
– multimers 100
– nonconjugative 393
poly(A) polymerase 236
PMF, *see* proton motive force 5
point mutation 53, 121
polyadenylation 236
polymyxin B 11
polypeptides
– circular 268
polyprotein 264
polyprotein processing 259, 266
porins 12, 13
Porphyromonas 264
postsegregational killing 80
postsynaptic step 89
post-transcriptional regulation 220
post-translational regulation 266
premutation 124, 131
pre-recruitment 201
presynapsis 86
presynaptic step 86
primases 56
primosomes 62
– preprimosome 62
– replication restart primosome 62
promoter 168, 171
promoter sequence 171
proofreading 174
prophage 29, 115, 399
protease 297
– ATP-dependent 298
protein cyclization 266
protein folding 280, 290
protein folding:disulfide bond 293
protein glycosylation 266
protein modification 266
protein remodeling 278
protein sorting signal 310

– cell wall sorting sequences 314
– lipoprotein sorting sequence 313
– signal sequences 310
– sorting motif 314
– stop-transfer sequences 312
– transmembrane sequence 312
protein splicing 266
protein-folding catalyst 292
proteolysis 191
Proteus 44
proton-motive force 4, 199, 327
pseudogenes 37
Pseudomonas 109
PYLIS 259
pyrrolysine 259

q
quorum sensing 214
quorum sensing system 215, 217

r
reactive oxygen species 138
rearrangement 121
RecA filament 87
recombination 83
– ectopic 83
– homologous 83
– illegitimate 112
– intermolecular 83
– intramolecular 83
– inversion systems 103
– postsynapsis step 89
– presynapsis step 85
– sequence-specific 97
– synapsis step 87
recombination machine 84
regulated intramembrane proteolysis 306
regulatory protein 172
regulon 362
relaxosome 390
replication 51, 148
– elongation 62
– bacterial chromosome 69
– direct rescue 149
– errors 125
– factory model 70
– fork 61, 92
– initiation 60, 61
– linear DNA 68
– mechanism 54
– recombination model 149
– restart 92
– termination 63
– train-on-track model 70
– unpaired lesions 92
replicative transposition 115
replisome 52, 93
repressosome 194
resolvase 98, 100, 103
resolvasome 89, 150
response regulator 206
resume codon 251
retroelement 44
retrohoming 239, 241
retron 44, 110
retrotransposition 240
reverse transcriptase 43, 110
Rhizobium 44
Rho factor 175, 238
Rhodobacter 166
Rhs element 41
riboregulation 222
ribosome recycling 235, 245, 253
riboswitch 220, 227, 231, 370
ribozyme 239
RIP, *see* regulated intramembrane proteolysis 306
RNA chaperone 242
– helicase 242
– polymerase 170
– polymerase: function of subunits 163
– splicing 239
– synthesis 162
RNAP, *see* DNA-dependent RNA polymerase 162
RNase III 233
RNase E 234
RNase G 235
rotations 23

s
SECIS 258
secretin 199, 200, 328, 339, 387
secretom 309
SEDS proteins 76
selenocysteine 258
sensor 341
sensor kinase 172
sequence-specific recombination 98
sequestration 66
Serratia 214
σ^{70} family 165
Shigella 31
shufflons 107
sigma factor 171
signal peptide peptidases 310
signal recognition particle 315, 316
signal sequence 310
silencer protein 155
single-strand break 62
single-stranded DNA 52
Sinorhizobium 205
site-specific recombination 97
S-layer 339
slow-stop mutant 54
SMC 71
sortases 314
SOS boxes 193
SOS response 143, 145
specifier sequence 187
starvation 341, 373
starvation stress 373
steric hindrance 189
sticky proteins 342

Streptococcus 382
Streptomyces 29
stress 172
– cold 349
– gene 341
– metabolic 371
– modulator 342
– osmotic 353
– oxidative 361
– pH 364
– response 227
– response pathway 341
– sensor 341
– starvation 373
stringent response 371, 373
supercoiling 31, 172
supercoils 56
superhelicity 32
– negative 32
symport 6
synapsis 87, 103
Syneccocystis 76
Synechococcus 350

t
TA module 79
telomer 67
temperature sensing 227
template DNA 62
termination of translocation 320
terminator 175
Thermotoga 59
Thermus 174
Tn*1000* 102
topoisomerase 32
toxin 224
toxin–antitoxin mechanism 79
toxin–antitoxin system 224
trans-translation 251, 252
transcription 170
– anti-termination 180
– arrested state 173
– attenuation 180–182
– elongation 173
– factor 172
– initiation 170
– termination factor (Rho) 175
– termination 175
– transcriptional pausing 173
– transcriptional slippage 174
transcriptional silencing 159
transcriptional units 36
transductants 400
transduction 381, 398
– generalized 400
– specialized 401
transferosome 390
transformation 381
– competence 382
transition 122
translation
– control 255
– elongation 248
– errors 244
– frameshifting 259
– initiation 247
– missence errors 244
– processivity errors 244
– regulator 242
– repression mechanisms 257
– repressor proteins 257
translational coupling 249
translesion synthesis 93, 149
translocase 316
translocation 309
– co-translational 319
– general secretion pathway 317
– post-translational 320
– protein sorting signal 309
– signal recognition particle 316
– sorting pathway 309
transport system 16
transposable element 113
transposase 113
transposing bacteriophage 113, 115
transposition 116
– conservative 116
– replicative 116
transposons 113, 114
trans-translation 252
transversion 122
TRAP 182, 221
trigger factor 272, 280
tubulin 76
turnover element 375
turgor 359
type III secretion pathway 330
type IV secretion pathway 333
type V secretion pathway 335

u
UP element 168
usher 24, 285
UV irradiation 140

v
very short patch repair 133
vesicle 336
Vibrio 37
vitamin sensing 228
VSR, *see* very short patch repair 133

y
Yersinia 286

z
Z-ring 77
zygotic induction 146